Value Distribution Theory of Meromorphic Functions

K. S. Charak
Department of Mathematics,
University of Jammu, Jammu 180 006, India.
E-mail: kscharak7@rediffmail.com

Abstract. The Value Distribution Theory of Nevanlinna is about a century old and still is an active area of research. It has a wide range of applications within and outside function theory. In this expository article we present this theory with history, recent developments and applications to Number Theory, Complex Dynamics, Factorization of Meromorphic Functions and Complex Differential Equations besides Value Sharing of Meromorphic Functions. Also, some open problems are presented for further investigations.

Keywords. Value distribution, Meromorphic function, Nevanlinna characteristic function, Deficient value.
2000 Mathematics Subject Classification: 30D35

1. Introduction

The roots of the equation such as

$$f(z) = a \tag{1.1}$$

where f is an entire function and a is a complex value, play an important role in solving certain theoretical or practical problems. It is especially important to investigate the number $n(r, a, f)$ of the roots of (1.1) and their distribution in a disk $|z| \leq r$, each root being counted with its proper multiplicity. It was the reserach on such topics that raised the curtains of the theory of value distribution of entire or meromorphic functions. The oldest result in the value distribution theory is the Fundamental Theorem of Algebra which appeared in Gauss's doctoral thesis in 1799. An easy consequence of it is the fact that a polynomial of degree n has n complex roots (counting with proper multiplicity). The entire function e^z behaves in entirely different manner. It omits the values zero and infinity entirely and every other value is assumed infinitely often. In the nineteenth century, the famous mathematician E. Picard obtained the path breaking result: *Any transcendental entire function f must take every finite complex value infinitely many times, with atmost one exception*. Later, E. Borel, introduced the concept of order of an entire function: *An entire function f is of order ρ, if for every $\epsilon > 0$, $\log^+ M(r, f) = O(r^{\rho+\epsilon})$ as $r \to \infty$, where $\log^+ |x|$ is the max*$\{\log |x|, 0\}$. By introduction of this concept, he gave the Picard's theorem a more precise formulation: *An entire function f of order $\rho(0 < \rho < \infty)$ satisfies*

$$\limsup_{r\to\infty} \frac{\log n(r, a, f)}{\log r} = \rho$$

for every finite complex value a, with atmost one exception. This result generally known as Picard–Borel Theorem, laid the foundation for the theory of value distribution and since then has been the source of many research accomplishments on this subject. The principal object and the tool of the theory were the class of entire functions and the maximum modulus, respectively. For meromorphic functions, maximum modulus could not be the proper tool to study their growth since it may become infinite in $|z| < r$ for finite values of r due to the presence of poles. It was Rolf Nevanlinna who in 1924 gave an ingenious interpretation to the well known Poisson–Jensen formula and got the study of the theory of meromorphic functions elevated to a new level by introducing a charcteristic function $T(r, f)$, now commonly known as Nevanlinna characteristic function for a meromorphic function f in a domain $|z| < R(R \leq \infty, 0 < r < R)$, as an efficient tool. What can be said about the distribution of the values of entire and meromorphic functions in general? This is the subject matter of Nevanlinna theory of value distribution which we shall refer to in this exposition as Nevanlinna Theory. In the present exposition we present a brief account-more or less a commentary- of

Nevanlinna Theory and its applications to variety of fields. For a more comprehensive treatment, the reader is referred to Hayman's famous book[30]. For further developments into different directions one can refer to the books by Cherry and Ye[13], Chuang and Yang[15], Gross[19], Lo[46], Nevanlinna[56], Ru[63], Rubel[64], and Zhang[81]. For historic development of the Nevanlinna Theory the reader is referred to [59].

2. The Poisson–Jensen Formula

Let $f(z)$ be a meromorphic function in $|z| \leq R$ $(0 < R < \infty)$, and let $a_j (j = 1, 2, \dots m)$ and $b_k (k = 1, 2, \dots n)$ be the zeros and poles of $f(z)$ in $|z| < R$ respectively, each zero and pole being counted according to their multiplicities. If $z = re^{i\theta}$ is a point in $|z| < R$ distinct from a_j and b_k, then

$$\log|f(z)| = \frac{1}{2\pi}\int_0^{2\pi} \log|f(Re^{i\phi})| \frac{R^2 - r^2}{R^2 - 2Rr\cos(\theta - \phi) + r^2} d\phi$$

$$+ \sum_{j=1}^{m} \log\left|\frac{R(z-a_j)}{R^2 - \bar{a}_j z}\right| - \sum_{k=1}^{n} \log\left|\frac{R(z-b_k)}{R^2 - \bar{b}_k z}\right|$$

Proof. Define

$$F(\zeta) = f(\zeta)\frac{\prod_{k=1}^{n}\left(\frac{R(\zeta - b_k)}{R^2 - \bar{b}_k\zeta}\right)}{\prod_{j=1}^{m}\left(\frac{R(\zeta - a_j)}{R^2 - \bar{a}_j\zeta}\right)} \tag{2.1}$$

Then $F(\zeta)$ has no zeros and poles in $|\zeta| \leq R$ and so it is analytic there. Choosing an analytic branch of $\log F(\zeta)$ in $|\zeta| \leq R$ and using Poisson's formula, we have

$$\log F(\zeta) = \frac{1}{2\pi}\int_0^{2\pi} \log F(Re^{i\phi}) \frac{R^2 - r^2}{R^2 - 2Rr\cos(\theta - \phi) + r^2} d\phi$$

Taking real parts and using $\Re\log F(\zeta) = \log|F(\zeta)|$, we have

$$\log|F(\zeta)| = \frac{1}{2\pi}\int_0^{2\pi} \log\left|F(Re^{i\phi})\right| \frac{R^2 - r^2}{R^2 - 2Rr\cos(\theta - \phi) + r^2} d\phi \tag{2.2}$$

From (2.1) and (2.2) we get,

$$\log|f(\zeta)| = \frac{1}{2\pi}\int_0^{2\pi} \log\left|F(Re^{i\phi})\right| \frac{R^2 - r^2}{R^2 - 2Rr\cos(\theta - \phi) + r^2} d\phi + \sum_{j=1}^{m} \log\left|\frac{R(\zeta - a_j)}{R^2 - \bar{a}_j\zeta}\right| - \sum_{k=1}^{n} \log\left|\frac{R(\zeta - b_k)}{R^2 - \bar{b}_k\zeta}\right| \tag{2.3}$$

Also, for $\zeta = Re^{i\theta}$ and $|a| < R$, we have

$$\left|\frac{R(\zeta - a)}{R^2 - \bar{a}\zeta}\right| = 1$$

implying that

$$\log\left|\frac{R(\zeta - a)}{R^2 - \bar{a}\zeta}\right| = 0$$

for $|\zeta| = R$ and so from (2.1) it follows that $\log|F(\zeta)| = \log|f(\zeta)|$ for $|\zeta| = R$. With this (2.3) reduces to

$$\log|f(z)| = \frac{1}{2\pi}\int_0^{2\pi} \log|f(Re^{i\phi})| \frac{R^2 - r^2}{R^2 - 2Rr\cos(\theta - \phi) + r^2} d\phi$$

$$+ \sum_{j=1}^{m} \log\left|\frac{R(z-a_j)}{R^2 - \bar{a}_j z}\right| - \sum_{k=1}^{n} \log\left|\frac{R(z-b_k)}{R^2 - \bar{b}_k z}\right|$$

as desired. □

Particularly important is the special case $z = 0$:

$$\log|f(0)| = \frac{1}{2\pi}\int_0^{2\pi} \log\left|f(Re^{i\phi})\right| d\phi + \sum_{j=1}^{m} \log\left|\frac{a_j}{R}\right| - \sum_{k=1}^{n} \log\left|\frac{b_k}{R}\right| \tag{2.4}$$

provided that $f(0) \neq 0, \infty$. Equation (2.4) is called *Jensen's formula*. If $f(0) = 0$ or ∞, then

$$f(z) = \sum_{k=m}^{\infty} C_k z^k,\ C_m \neq 0,\ m \in \mathbb{Z}.$$

In fact, $m > 0$ if the origin is a zero of order m, and $m < 0$ if the origin is a pole of order m. Then $g(0) \neq 0, \infty$, and has same zeros and poles as $f(z)$ in $o < |z| < R$. Now applying Jensen's formula to $g(z)$, we get

$$\log|C_m| = \frac{1}{2\pi}\int_0^{2\pi} \log|f(Re^{i\phi})| d\phi + \sum_{j=1}^{m} \log\left|\frac{a_j}{R}\right| - \sum_{k=1}^{n} \log\left|\frac{b_k}{R}\right| - m\log R \tag{2.5}$$

In Nevanlinna's hands Jensen's formula, established by Jensen in 1899, was to lead to amazing consequences.

3. Reformulation of Jensen's Formula-The Birth of Nevanlinna Theory

For non-negative real number x, define

$$\log^+ x = max\{0, \log x\}.$$

Then one can easily find that $\log^+$ has the following properties:

1. $\log x \leq \log^+ x$
2. $\log^+ x \leq \log^+ y$ for $x \leq y$
3. $\log x = \log^+ x - \log^+ \frac{1}{x}$
4. $\log|x| = \log^+ x - \log^+ \frac{1}{x}$
5. $\log^+(\prod_{k=1}^{n} x_k) \leq \sum_{k=1}^{n} \log^+ x_k$
6. $\log^+(\sum_{k=1}^{n} x_k) \leq \log n + \sum_{k=1}^{n} \log^+ x_k$

For a meromorphic function f in $|z| \leq R$ we denote by $n(t, f)$ the number of poles of f counting multiplicities in $|z| \leq t$, $0 < t < R$; by $n(t, \frac{1}{f-a})$ or $n(t, a)$ the number of a-points of f in $|z| \leq t$, counting multiplicities. For $a \neq \infty$, the proximity function is given by

$$m\left(R, \frac{1}{f-a}\right) = m(R, a) = \frac{1}{2\pi}\int_0^{2\pi} \log^+ \frac{1}{\left|f(Re^{i\phi}) - a\right|} d\phi$$

and

$$m(R, f) = m(R, \infty) = \frac{1}{2\pi}\int_0^{2\pi} \log^+ \left|f(Re^{i\phi})\right| d\phi$$

The counting function of a-points of f is defined as

$$N\left(R, \frac{1}{f-a}\right) = N(R, a) = \int_0^R \frac{n(t, a) - n(0, a)}{t} dt + n(0, a)\log R$$

And

$$N(R, f) = N(R, \infty) = \int_0^R \frac{n(t, f) - n(0, f}{t} dt + n(0, f)\log R$$

$m(R,a)$ defined as the mean value of $\log^+\left|\frac{1}{f-a}\right|$ (or $\log^+|f|$ if $a=\infty$) on the circle $|z|=R$ receives a significant contribution only from the arcs on the circle where the function value differs very little from the given value a. The magnitude of the proximity function $m(R,a)$ can thus be considered as a measure for the mean deviation on the circle $|z|=R$ of the function f from the value a.

The counting function of a-points $N(R,a)$ indicates how densely the roots of the equation $f(z)=a$ are distributed in the mean in the disk $|z|<R$. The larger the number of a-points, the faster the counting function grows with R.

Now let us go back to (2.4). By using third property of $\log^+$, we find that

$$\frac{1}{2\pi}\int_0^{2\pi}\log\left|f(Re^{i\phi})\right|d\phi=\frac{1}{2\pi}\int_0^{2\pi}\log^+\left|f(Re^{i\phi})\right|d\phi-\frac{1}{2\pi}\int_0^{2\pi}\log^+\frac{1}{\left|f(Re^{i\phi})\right|}d\phi$$
$$=m(R,f)-m\left(R,\frac{1}{f}\right) \tag{3.1}$$

Next, let $|b_k|=r_k$. Then by Riemann-Stieltjes integral, we have

$$\sum_{k=1}^{n}\log\frac{R}{r_k}=\int_0^R\log\frac{R}{t}d[n(t,f)-n(0,f)]$$
$$=[n(t,f)-n(0,f)]\log\frac{R}{t}|_0^R+\int_0^R\frac{n(t,f)-n(0,f)}{t}dt$$
$$=\int_0^R\frac{n(t,f)-n(0,f)}{t}dt$$

Similarly, we have

$$\sum_{k=1}^{n}\log\frac{R}{|a_j|}=\int_0^R\frac{n\left(t,\frac{1}{f}\right)-n\left(0,\frac{1}{f}\right)}{t}dt$$

With these notations and definitions (2.4) and (2.5)(with $m=n(0,\frac{1}{f})-n(0,f)$) become

$$m(R,f)-m\left(R,\frac{1}{f}\right)-N\left(R,\frac{1}{f}\right)+N(R,f)=\log|f(0)|, \tag{3.2}$$

and

$$m(R,f)-m\left(R,\frac{1}{f}\right)-N\left(R,\frac{1}{f}\right)+N(R,f)=\log|C_m| \tag{3.3}$$

respectively. Define

$$T(R,f)=m(R,f)+N(R,f) \tag{3.4}$$

which is called the *Nevanlinna Characteristic function of* f and after some more discussion we shall show that $T(R,f)$ is an increasing convex function of $\log R$ and behaves like $\log M(R,f)$ when f happens to be an entire function, where $M(R,f)=max_{|z|=R}|f(z)|$.

With (3.4), (3.2) and (3.3) respectively reduce to

$$T\left(R,\frac{1}{f}\right)=T(R,f)-\log|f(0)|$$

and

$$T\left(R,\frac{1}{f}\right)=T(R,f)-\log|C_m|$$

Thus in general, we have,

$$T\left(R, \frac{1}{f}\right) = T(R, f) + O(1), \tag{3.5}$$

where $O(1)$ is a bounded term. Thus Jensen's formula says that $T(R, f)$ and $T\left(R, \frac{1}{f}\right)$ differ only by a bounded term. If we look at the physical meaning of proximity and counting functions we find that $T(R, f)$ is the total affinity of the function f for the value ∞ in $|z| \le R$ whereas $T\left(R, \frac{1}{f}\right)$ is the total affinity of the function f for the value 0 in $|z| \le R$. Thus Jensen's formula says that the total affinity of f for 0 and ∞ is of the same order. Not only this, Nevanlinna proved that this is true for any value $a \in \mathbb{C}$-the fact known as the First Fundamental Theorem of Nevanlinna.

4. Properties of Characteristic Function and First Fundamental Theorem of Nevanlinna

Using properties 5 and 6 of $\log^+$ one can easily deduce that if f_k, $(k = 1, 2, \ldots n)$ are meromorphic functions in $|z| < R$, then for $0 < r < R$

1. $m\left(r, \prod_{k=1}^{n} f_k\right) \le \sum_{k=1}^{n} m(r, f_k)$
2. $m\left(r, \sum_{k=1}^{n} f_k\right) \le \log n + \sum_{k=1}^{n} m(r, f_k)$

Also, since the order of the pole of $\sum_{k=1}^{n} f_k$ at z_0 does not exceed the sum of the orders of the poles of f_k at z_0, we have

$$N\left(r, \sum_{k=1}^{n} f_k\right) \le \sum_{k=1}^{n} N(r, f_k)$$

This together with property 2 above, gives

$$T\left(r, \sum_{k=1}^{n} f_k\right) \le \log n + \sum_{k=1}^{n} T(r, f_k). \tag{4.1}$$

Similarly, we have

$$T\left(r, \prod_{k=1}^{n} f_k\right) \le \sum_{k=1}^{n} T(r, f_k). \tag{4.2}$$

The most important property of the Characteristic function which makes it a suitable yard stick for measuring the growth of meromorphic functions including entire functions as a sepecial case is

- $T(r, f)$ is an increasing function of r and a convex function of $\log r$

To establish this property we first find an alternative representation of $T(r, f)$.

Applying Jensen's formula with $R = 1$ to the function $g(z) = a - z$, we get

$$\frac{1}{2\pi} \int_0^{2\pi} \log^+ |Re^{i\phi} - a| d\phi = \log^+ |a|, \ \forall a \in \mathbb{C}. \tag{4.3}$$

Let $0 < r < R$. Applying Jensen's formula to the function $f(z) - e^{i\theta}$, we get

$$\log|f(0) - e^{i\theta}| = \frac{1}{2\pi} \int_0^{2\pi} \log^+ |f(re^{i\phi} - e^{i\theta}| d\phi - N\left(r, \frac{1}{f - e^{i\theta}}\right) + N(r, f - e^{i\theta})$$

Integrating w.r.t. θ from 0 to 2π, we obtain

$$\frac{1}{2\pi} \int_0^{2\pi} \log|f(0) - e^{i\theta}| d\theta = \frac{1}{2\pi} \int_0^{2\pi} \left[\frac{1}{2\pi} \int_0^{2\pi} \log^+ |f(re^{i\phi}) - e^{i\theta}| d\phi\right] d\theta$$

$$- \frac{1}{2\pi} \int_0^{2\pi} N\left(r, \frac{1}{f - e^{i\theta}}\right) d\theta + \frac{1}{2\pi} \int_0^{2\pi} N(r, f - e^{i\theta}) d\theta$$

Using (4.3) with $a = f(re^{i\phi})$ we find that

$$\begin{aligned}\frac{1}{2\pi}\int_0^{2\pi} \log|f(0) - e^{i\theta}|d\theta &= \frac{1}{2\pi}\int_0^{2\pi} \log^+|f(re^{i\phi})|d\phi \\ &\quad - \frac{1}{2\pi}\int_0^{2\pi} N\left(r, \frac{1}{f - e^{i\theta}}\right)d\theta + N(r, f) \\ &= m(r, f) + N(r, f) - \frac{1}{2\pi}\int_0^{2\pi} N\left(r, \frac{1}{f - e^{i\theta}}\right)d\theta \\ &= T(r, f) - \frac{1}{2\pi}\int_0^{2\pi} N\left(r, \frac{1}{f - e^{i\theta}}\right)d\theta\end{aligned}$$

This by another application of (4.3) yields that

$$T(r, f) = \frac{1}{2\pi}\int_0^{2\pi} N\left(r, \frac{1}{f - e^{i\theta}}\right)d\theta + \log^+|f(0)|. \tag{4.4}$$

Equation (4.4) is called *Henri Cartan's Identity* and this is the alternative representation of $T(r, f)$.

Since $N(r, e^{i\theta})$ is a non-decreasing function of r, Cartan's identity (4.4) yields that $T(r, f)$ is an increasing function of r. Again from Cartan's identity, we have

$$\frac{dT(r, f)}{d\log r} = \frac{d}{d\log r}\frac{1}{2\pi}\int_0^{2\pi}\left(\int_0^r n(t, e^{i\theta})\frac{dt}{t}\right)d\theta = \frac{1}{2\pi}\int_0^{2\pi} n(r, e^{i\theta})d\theta.$$

Since the right hand side is non-negative and non-decreasing with r, it follows that $T(r, f)$ is a convex function of $\log r$. Thus $T(r, f)$ is an increasing function of r and a convex function of $\log r$ as desired.

- $m(r, a)$ is neither increasing nor decreasing in general.

For example, consider

$$f(z) = \frac{z}{1 - z^2}.$$

Then $|f(z)| < 1$ for $|z| < \frac{1}{2}$ and $|z| > 2$. This implies that $m(r, f) = 0$ for $r \leq \frac{1}{2}$ and $r \geq 2$. On the other for $r = 1$ we have

$$\begin{aligned}m(r, f) &= \frac{1}{2\pi}\int_0^{2\pi} \log^+\left|f(e^{i\phi})\right|d\phi \\ &= \frac{1}{2\pi}\int_0^{2\pi} \log^+\frac{1}{\left|1 - e^{2i\phi}\right|}d\phi \to \infty \text{as} \phi \to 0\end{aligned}$$

and so $m(r, f) > 0$, infact it is large enough.

Theorem 1 (First Fundamental Theorem of Nevanlinna). *Let $f(z)$ be a non-constant meromorphic function in $|z| < R(0 < R \leq \infty)$. Then for any finite complex number a, we have*

$$T\left(r, \frac{1}{f - a}\right) = T(r, f) - \log|C_m| + \epsilon(r, a)$$

where $|\epsilon(r, a)| \leq \log^+|a| + \log 2$, $0 < r < R$, and C_m is the first non-zero coefficient in the Laurent series expansion of $f - a$ about the origin.

Proof. Define $h(z) = f(z) - a$. Then $N(r, h) = N(r, f)$. Since

$$\log^+|h| = \log^+|f - a| \leq \log^+|f| + \log^+|a| + \log 2$$

and

$$\log^+|f| = \log^+|f-a+a| \le \log^+|f-a| + \log^+|a| + \log 2,$$

integrating these inequalities, we have

$$m(r,h) \le m(r,f) + \log^+|a| + \log 2$$

and

$$m(r,f) \le m(r,h) + \log^+|a| + \log 2.$$

Thus,

$$\epsilon(r,a) := m(r,h) - m(r,f)$$

satisfies that

$$|\epsilon(r,a)| \le \log^+|a| + \log 2.$$

Now applying Jensen's formula to h we get

$$\begin{aligned} T\left(r,\frac{1}{h}\right) &= T(r,h) - \log|C_m| \\ &= m(r,f) + N(r,f) - \log|C_m| + \epsilon(r,a) \\ &= T(r,f) - \log|C_m| + \epsilon(r,a). \end{aligned}$$

□

Remark 1. *The last Theorem can also be put as*

$$T(r,f) = m(r,a) + N(r,a) + O(1), \tag{4.5}$$

where $O(1)$ remains bounded as $r \to \infty$.

Corollary 1. $\frac{1}{2\pi}\int_0^{2\pi} m(r,e^{i\theta})d\theta \le \log 2.$

Proof. By First Fundamental Theorem of Nevanlinna we have

$$T(r,f) = m(r,e^{i\theta}) + N(r,e^{i\theta}) + \log^+\left|f(0)-e^{i\theta}\right| + \epsilon(r,e^{i\theta})$$

where $\left|\epsilon(r,e^{i\theta})\right| \le \log 2$.

Integrating on both sides with respect to θ from 0 to 2π and using Cartan's identity, we get

$$T(r,f) = \frac{1}{2\pi}\int_0^{2\pi} m(r,e^{i\theta})d\theta + T(r,f) + \frac{1}{2\pi}\int_0^{2\pi} \epsilon(r,e^{i\theta})d\theta$$

which implies that

$$\begin{aligned} \frac{1}{2\pi}\int_0^{2\pi} m(r,e^{i\theta})d\theta &= -\frac{1}{2\pi}\int_0^{2\pi} \epsilon(r,e^{i\theta})d\theta \\ &\le \log 2. \end{aligned}$$

□

Note: The last result shows that $m(r,a)$ is bounded in the mean on the circle $|a| = 1$. Thus if $T(r,f)$ is large, $m(r,a)$ bounded and $N(r,a)$ is nearly equal to $T(r,f)$ for most values of a, in certain sense.

As mentioned earlier, if f is an entire function, then $T(r,f)$ behaves as $\log^+ M(r,f)$ and this is a consequence of the following fundamental inequality:

Theorem 2. *If f is holomorphic for $|z| \leq R$ and $M(r, f) = max\{|f(z)| : |z| \leq r\}$, then*

$$T(r, f) \leq \log^+ M(r, f) \leq \frac{R+r}{R-r} T(R, f)$$

$(0 \leq r < R)$.

Proof. Since f is holomorphic in $|z| \leq R$, for $0 \leq r < R$, we have

$$\begin{aligned}
T(r, f) = m(r, f) &= \frac{1}{2\pi} \int_0^{2\pi} \log^+ \left| f(re^{i\theta}) \right| d\theta \\
&\leq \frac{1}{2\pi} \int_0^{2\pi} \log^+ M(r, f) d\theta \\
&= \log^+ M(r, f)
\end{aligned}$$

which proves the left hand inequality. The right hand inequality trivially holds if $M(r, f) \leq 1$. Thus we assume that $M(r, f) > 1$. Let z_0 be a point on the circle $|z| = r$ such that $|f(z_0)| = M(r, f)$. Since f has no poles in $|z| < R$ and $\left| \frac{R(z-a_j)}{R^2 - \bar{a}_j z} \right| < 1$, Poisson–Jensen formula yields that

$$\begin{aligned}
\log^+ M(r, f) = \log |f(z_0)| &= \frac{1}{2\pi} \int_0^{2\pi} \log^+ \left| f(Re^{i\phi}) \right| \frac{R^2 - r^2}{R^2 - 2Rr\cos(\theta - \phi) + r^2} d\phi \\
&\leq \frac{R+r}{R-r} \frac{1}{2\pi} \int_0^{2\pi} \log^+ \left| f(Re^{i\phi}) \right| = \frac{R+r}{R-r} m(R, f) \\
&= \frac{R+r}{R-r} T(R, f).
\end{aligned}$$

□

Since by last theorem $T(r, f) \sim \log^+ M(r, f)$ as $r \to \infty$, the order ρ of a meromorphic function f is defined as

$$\rho = \limsup_{r \to \infty} \frac{\log^+ T(r, f)}{\log r}.$$

5. Examples

1. Consider the rational function

$$f(z) = \frac{P(z)}{Q(z)} = \frac{a_n z^n + a_{n-1} z^{n-1} + \ldots + a_0}{b_m z^m + b_{m-1} z^{m-1} + \ldots + b_0}, \ a_n, \ b_m \neq 0.$$

Distinguish the following two cases:

Case 1. When $m \geq n$. In this case $\lim_{|z| \to \infty} f(z)$ is finite, so there is a positive real number r_0 such that $n(r, f) = m \ \forall r \geq r_0$. Thus

$$\begin{aligned}
N(r, f) &= \int_0^{r_0} \frac{n(t, \infty) - n(0, \infty)}{t} dt + \int_{r_0}^{r} \frac{m - n(0, \infty)}{t} dt + n(0, \infty) \log r \\
&= (m - n(0, \infty))(\log r - \log r_0) + n(0, \infty) \log r + O(1) \\
&= m \log r - m \log r_0 + n(o, \infty) \log r_0 + O(1) = m \log r + O(1).
\end{aligned}$$

Next, note that for polynomial $P(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_0$ with $a_n \neq 0$, given positive ϵ there is an $r_0 > 0$ such that $\forall r = |z| > r_0$ we have

$$(1 - \epsilon)|a_n| r^n \leq |P(z)| \leq (1 + \epsilon)|a_n| r^n.$$

Thus $\forall r \geq r_0$ we can assume that $|P(z)| = |a_n|r^n(1+o(1))$ and $|Q(z)| = |b_m|r^m(1+o(1))$. This implies that $\log^+|f| = O(1)$ and so $m(r, f) = O(1)$. Hence in this case

$$T(r, f) = m \log r + O(1) = O(\log r)$$

Case 2. When $m < n$. In this case we apply Jensen formula and the arguments used in *Case 1* and get

$$T(r, f) = T\left(r, \frac{1}{f}\right) + O(1) = n \log r + O(1) = O(\log r).$$

Thus for a rational function f we have $T(r, f) = O(\log r)$. Also, the converse of this statement holds. That is, if f is a meromorphic function with $T(r, f) = O(\log r)$, then f is a rational function.

2. Let $f(z) = e^z$. Then for $a = \infty$ we have

$$\begin{aligned} m(r, f) &= \frac{1}{2\pi}\int_o^{2\pi} \log^+|e^{re^{i\phi}}|d\phi \\ &= \frac{1}{2\pi}\int_o^{2\pi} \log^+ e^{rcos\phi}d\phi \\ &= \frac{1}{2\pi}\int_{\frac{-\pi}{2}}^{\frac{\pi}{2}} rcos\phi d\phi \\ &= \frac{r}{\pi} \end{aligned}$$

Since e^z is entire, $N(r, f) = 0$ and so

$$T(r, f) = \frac{r}{\pi}.$$

Also, if $a = 0$ we have

$$m(r, a) = \frac{r}{\pi}, \quad N(r, a) = 0.$$

Further, if $a \neq 0, \infty$ and if z_0 is a root of the equation $f(z) = a$, then by periodicity of f, other roots of this equation are of the form $z_0 + 2k\pi i, \ k \in \mathbb{Z}$ and hence the number of roots of $f(z) = a$ in $|z| \leq t$ is

$$n(t, a) = \frac{t}{\pi} + O(1) \ \Rightarrow N(r, a) = \frac{r}{\pi} + O(\log r)$$

Also, after doing some computations one gets that $m(r, a) = O(1)$.

It may be remarked here that after finding $T(r, f)$ we appeal to the First Fundamental Theorem of Nevanlinna to obtain

$$T\left(r, \frac{1}{f-a}\right) = T(r, f) + O(1) \text{ and this implies that } N(r, a) = \frac{r}{\pi} + O(\log r).$$

3. Let $P(z) = \sum_{k=0}^{n} a_k z^k$ be a polynomial of degree n and consider the function $f(z) = e^{P(z)}$. First we calculate $T(r, f)$ when $P(z) = a_n z^n$.

Let $a_n = |a_n|e^{i\phi}, \ z = re^{i\theta}$. Then

$$|f(z)| = e^{|a_n|r^n cos(n\theta+\phi)}$$

and so we have

$$
\begin{aligned}
m(r, f) &= \frac{1}{2\pi}\int_{o}^{2\pi} \log^{+} e^{|a_n|r^n cos(n\theta+\phi)} d\theta \\
&= \frac{1}{2\pi}\int_{\phi}^{2n\pi+\phi} \log^{+} e^{|a_n|r^n cos\eta} d\eta \\
&= \frac{1}{2n\pi}\int_{0}^{2n\pi} \log^{+} e^{|a_n|r^n cos\eta} d\eta \text{ (by periodicity)} \\
&= \frac{|a_n|r^n}{2\pi}\int_{\frac{-\pi}{2}}^{\frac{\pi}{2}} cos\eta d\eta \\
&= \frac{|a_n|r^n}{\pi}
\end{aligned}
$$

Since f is entire, therefore,

$$
T(r, f) = m(r, f) = \frac{|a_n|r^n}{\pi}.
$$

Now Since

$$
T(r, f) = T(r, e^{P(z)}) = T(r, e^{a_n z^n} \cdot e^{a_{n-1} z^{n-1}} \cdots e^{a_0})
$$

and

$$
T(r, e^{a_{n-k} z^{n-k}}) = \frac{|a_{n-k}|}{\pi} r^{n-k} = o\left(\frac{|a_n|}{\pi} r^n\right) = o(T(r, e^{a_n z^n}),
$$

it follows that

$$
T(r, f) = T(r, e^{P(z)}) \sim T(r, e^{a_n z^n}) = \frac{|a_n|}{\pi} r^n \ (r \to \infty).
$$

4. Let f be a non-constant meromorphic function and consider the function

$$
g = \frac{af + b}{cf + d}
$$

where a, b, c, and d are constants with $ad - bc \neq 0$.

For $c = 0$, inequality (4.1) clearly implies that

$$
T(r, g) = T(r, f) + O(1).
$$

so, let us assume that $c \neq 0$. Define

$$
g_1 = f + \frac{d}{c},
$$

$$
g_2 = cg_1,
$$

$$
g_3 = \frac{1}{g_2},
$$

and

$$
g_4 = \frac{(bc - ad)}{c} g_3.
$$

Then

$$g = g_4 + \frac{a}{c}$$

Now by using inequalities (4.1) and (4.2) and First Fundamental Theorem of Nevanlinna, we get

$$\begin{aligned} T(r, g) &= T(r, g_4) + O(1) \\ &= T(r, g_3) + O(1) \\ &= T(r, g_2) + O(1) \\ &= T(r, g_1) + O(1) \\ &= T(r, f) + O(1) \end{aligned}$$

6. Second Fundamental Theorem of Nevanlinna

Relation (4.5) of the First Fundamental Theorem of Nevanlinna gave rise to a question of the relative size of the components $m(r, a)$ and $N(r, a)$ in the invariant sum $m(r, a) + N(r, a)$. The answer to this question was given by Nevanlinna with great accuracy in July 1924 (see [52]) by proving the inequality

$$T(r, f) < N(r, a) + N(r, b) + N(r, c) + S(r, f), \tag{6.1}$$

where a, b, c are three distinct values and $S(r, f)$ is in general small compared to $T(r, f)$. This inequality shows that in relation (4.5) the term $N(r, a)$ is usually larger than $m(r, a)$. For example, if $\limsup_{r\to\infty} \frac{N(r,a)}{T(r,f)} = 0$ for two values of a, then $\limsup_{r\to\infty} \frac{N(r,a)}{T(r,f)} = 1$ for all other values of a.

In 1925 Nevanlinna[53] presented a large survey of his theory of meromorphic functions which is regarded as Nevanlinna's main work. In 1943 H. Weyl[74] wrote about it: "The appearance of this paper has been one of the few great mathematical events of our century". After going through the observations made by Littlewood and Collingwood in 1924 that the method of the proof of inequality (6.1) can actually be applied to arbitrary many values instead of just three values, Nevanlinna proved his second fundamental theorem as

Theorem 3. *Let f be a non-constant meromorphic function in $|z| \le r$, and let $a_1, a_2, \ldots, a_q (q \ge 3)$ be distinct finite complex numbers. Then*

$$m(r, f) + \sum_{k=1}^{q} m(r, a_k) \le 2T(r, f) - N_1(r, f) + S(r, f),$$

where $N_1(r, f)$ is positive and is given by

$$N_1(r, f) = N\left(r, \frac{1}{f'}\right) + 2N(r, f) - N(r, f'), \tag{6.2}$$

and

$$S(r, f) = m\left(r, \frac{f}{f'}\right) + m\left(r, \sum_{k=1}^{q} \frac{f'}{(f - a_k)}\right) + q \log^+ \frac{3q}{\delta} + \log 2 + \log \frac{1}{|f'(0)|}, \tag{6.3}$$

with modifications if $f(0) = 0$ or ∞ or $f'(0) = 0$.

The quantity $S(r, f)$ given by (6.3) in Theorem 3 plays in general(as we shall see in the following section)the role of an unimportant error term. Theorem 3 with this fact is what is known as the second fundamental theorem of Nevanlinna. Thus the second fundamental theorem of Nevanlinna says that the sum of any number of terms $m(r, a_k)$ can not in general be much greater than $2T(r, f)$. Just not to make the exposition too bulky and keeping in view the simplicity of the proof of Theorem 3 we refer the reader to [30].

6.1 Estimation of $S(r, f)$

The most interesting and of course challenging question in the Nevanlinna Theory is the sharp estimation of the term $S(r, f)$. Following are some of the known estimates of $S(r, f)$ (see [30]):

Theorem 4. *Let f be a non-constant meromorphic function in $|z| < R \leq +\infty$. Then*

1. *if $R = +\infty$,*

$$S(r, f) = O\{\log T(r, f)\} + O(\log r)$$

   as $r \to \infty$ through all values if f is of finite order, and as $r \to \infty$ outside a set E of finite linear measure otherwise.
2. *if $0 < R < +\infty$,*

$$S(r, f) = O\left\{\log^+ T(r, f) + \log \frac{1}{R - r}\right\}$$

   as $r \to \infty$ outside a set E with $\int_E \frac{dr}{R-r} < +\infty$.

As an immediate deduction from Theorem 4, we have

Theorem 5. *Let f be a non-constant meromorphic function in $|z| < R \leq +\infty$. Then*

$$\frac{S(r, f)}{T(r, f)} \to 0, \text{ as } r \to R, \tag{6.4}$$

with the following provisos:

1. *If $R = +\infty$, and if f is of finite order, then (6.4) holds without any restriction.*
2. *If $R = +\infty$, and f is of infinite order, then (6.4) holds as $r \to \infty$ outside a set E of finite length.*
3. *If $R = +\infty$, and*

$$\liminf_{r \to R} \frac{T(r, f)}{\log\left\{\frac{1}{R-r}\right\}} = +\infty$$

   then (6.4) holds as $r \to R$ through a suitable sequence of values of r.

In late eightees S. Lang[38] after pointing out that Vojta[73] had observed several connections between Number Theory and Nevanlinna Theory raised the problem of finding the best possible form of the upper bound for $S(r, f)$. After S. Lang and W. Cherry[39], P. Wong[75] and Z. Ye[79], A. Hinkkanen[33] obtained a (better) sharp upper bound for $S(r, f)$: *Let f be a meromorphic function in the plane. Then for any positive increasing functions $\frac{\phi(t)}{t}$ and $p(t)$ with $\int_1^\infty \frac{dt}{\phi(t)} < \infty$ and $\int_1^\infty \frac{dt}{p(t)} = \infty$,*

$$S(r, f) \leq \log^+ \left\{\frac{\phi(T(r, f))}{p(r)}\right\} + O(1), \tag{6.5}$$

as $r \to \infty$ outside a set E with $\int_E \frac{dr}{p(r)} < \infty$. Further, if $\frac{\psi(t)}{t}$ is positive and increasing and $\int_1^\infty \frac{dt}{\psi(t)} = \infty$, then there is an entire function f such that

$$S(r, f) \geq \log \psi(T(r, f))$$

outside a set of finite linear measure. A. Hinkkanen[33] has also obtained analogous results for functions meromorphic in a disk. For further recent developments about the connections between Number Theory and Nevanlinna Theory, one can refer to Cherry and Ye[13] and Ru[63].

In the Second Fundamental Theorem of Nevanlinna taking $S(r, f) = o(T(r, f))$ as $r \to \infty$, for some set $E \subset [0, \infty)$ of finite measure, we have

$$m(r, f) + \sum_{k=1}^{q} m(r, a_k) \leq (2 + o(1))T(r, f) - N_1(r, f)$$

Adding $\sum_{k=1}^{q} N(r, a_k) + N(r, f)$ on both sides and using the First Fundamental Theorem of Nevanlinna, we get

$$(q - 1 - o(1))T(r, f) \leq \sum_{k=1}^{q} N(r, a_k) + N(r, f) - N_1(r, f),\ r \notin E.$$

Since $N_1(r, f)$ counts the multiple points, it follows from the last inequality that

$$(q - 1 - o(1))T(r, f) \leq \sum_{k=1}^{q} \overline{N}(r, a_k) + \overline{N}(r, f),\ r \notin E,$$

where $\overline{N}(r, a) = \int_0^r \frac{\overline{n}(t,a)dt}{t}$ wherein $\overline{n}(t, a)$ denotes the number of a-points of f ignoring multiplicity in the disk $|z| \leq r$. The last two inequalities can further be written as

$$(q - 2 - o(1))T(r, f) \leq \sum_{k=1}^{q} N(r, a_k) - N_1(r, f),\ r \notin E. \tag{6.6}$$

and

$$(q - 1 - o(1))T(r, f) \leq \sum_{k=1}^{q} \overline{N}(r, a_k),\ r \notin E. \tag{6.7}$$

if $a_1, a_2 \dots a_q \in \overline{\mathbb{C}}$ are distinct.

Now suppose that a transcendental meromorphic function f that takes three distinct values $a_1, a_2, a_3 \in \overline{\mathbb{C}}$ only finitely many times. Then $N(r, a_k) = O(\log r)$ for $k = 1, 2, 3$. Now inequality (6.6) with $q = 3$ yields that $(1 - o(1))T(r, f) = O(\log r)$ which shows that f is rational, a contradiction. This is nothing but another proof of Picard's Little Theorem using Nevanlinna Theory. This is most elementary proof of Picard's Little Theorem as compared to its original proof using elliptic modular functions. Also, this proof has an advantage that it gets generalized to prove higher dimensional analogs of Picard's Theorem. We may view the Second Fundamental Theorem of Nevanlinna and the preceding inequalities deduced from it as a quantitative version of Picard's Theorem.

Definition 1. *A value $a \in \overline{\mathbb{C}}$ is said to be a perfectly branched value of a meromorphic function f if only finitely many of the zeros of $f(z) - a$ are simple.*

The Nevanlinna Theory also gives information about the multiplicities with which the values are assumed by meromorphic functions.

Theorem 6. *A transcendental meromorphic function has at most four perfectly branched values.*

Proof. Suppose $f(z) - a$ has q simple zeros at $a_k,\ k = 1, 2 \dots q$. Then

$$\overline{N}(r, a_k) \leq \frac{1}{2} N(r, a_k) + O(\log r),$$

and so inequality (6.7) yields

$$\begin{aligned}(q - 2 - o(1))T(r, f) &\leq \sum_{k=1}^{q} \overline{N}(r, a_k) \\ &\leq \frac{1}{2} \sum_{k=1}^{q} N(r, a_k) + O(\log r) \\ &\leq \left(\frac{q}{2} + o(1)\right) T(r, f),\ r \notin E.\end{aligned}$$

This implies that $q - 2 \leq \frac{q}{2}$ and thus $q \leq 4$. □

By similar arguments we obtain that a transcendental entire function has at most two finite perfectly branched values. The numbers 4 and 2 here are sharp. For example, the Weierastrass $\wp$-function has four perfectly branched values and the *sine* function has two perfectly branched values.

6.3 An Extension of the Second Fundamental Theorem

R. Nevanlinna[54] asked if the complex numbers a_k in the Second Fundamental Theorem can be replaced by functions of slower growth as compared to f. He himself proved it for $k = 3$ (for proof see [30] p-47). Chi-tai Chuang[14] obtained the first non-trivial result.

Theorem 7. *Let f and $\phi_k(k = 1, 2, \dots, q)$ be meromorphic functions on $\mathbb{C}$ such that ϕ_k,s are distinct and $T(r, \phi_k) = o(T(r, f))$ for $k = 1, 2 \dots, q$. Then*

$$(q - 1 - o(1))T(r, f) < \sum_{k=1}^{q} N(r, \phi_k) + q\overline{N}(r, f) + S(r, f)$$

where $S(r, f) = o(1)$ as $r \to \infty$, for some set $E \subset [0, \infty)$ of finite measure.

The complete solution of Nevanlinna's problem was given by Osgood[60] and Steinmetz[70], independent of one another.

Theorem 8. *Let f and $\phi_k(k = 1, 2, \dots, q)$ be meromorphic functions on $\mathbb{C}$ such that ϕ_k, s are distinct and $T(r, \phi_k) = o(T(r, f))$ for $k = 1, 2 \dots, q$. If ϵ is a positive number, then*

$$m(r, f) + \sum_{k=1}^{q} m(r, \phi_k) \leq (2 + \epsilon)T(r, f),$$

except on a set E_ϵ of r with finite linear measure.

The functions ϕ_k, s in the last two theorems are generally known as small functions of f excepting W. Cherry and Z. Ye[13] who call such functions as slowly moving targets to fit it into the language of Number Theory. We denote by $n^{(k)}(r, a)$ the number of a-points of f in $|z| \leq r$ of maximum multiplicity k and thus we define the truncated counting function $N^{(k)}(r, a)$ as

$$N^{(k)}(r, a) = \int_0^r \frac{n^{(k)}(t, a)dt}{t}.$$

The above theorems are called versions of the Second Fundamental Theorem for moving targets. The Nevanlinna's extension to three small functions is described by Cherry and Ye as *three moving target theorem of Nevanlinna* stated by using truncated counting function and contains a *ramification term*. Finding a proof for moving target second fundamental theorem that works with truncated counting functions is an important open problem.

7. Applications of Nevanlinna Theory

In this section one can realise the power of Nevanlinna Theory. Nevanlinna Theory has a wide range of applications starting from Number theory to Probability and Statistics and to Theoretical Physics. Here we shall touch upon a few areas within Function Theory and of course Number Theory.

7.1 Defect Relation-A Weaker Reformulation of the Second Fundamental Theorem

The reformulation of the Second Fundamental Theorem we are about to take up, is called the Defect Relation. It is infact a weaker reformulation of the Second Fundamental Theorem of Nevanlinna but has some nice consequences. For this, we first introduce the term *defect*. Let f be a non-constant meromorphic function on $\mathbb{C}$. Then for $a \in \overline{\mathbb{C}}$, we define the terms $\delta(a, f)$, $\theta(a, f)$ and $\Theta(a, f)$ known as Nevanlinna defect or deficiency of a w.r.t. f, the index of multiplicity or the ramification defect and the truncated defect of order 1; as

$$\delta(a, f) = 1 - \limsup_{r\to\infty} \frac{N(r, a)}{T(r, f)}$$

$$\theta(a, f) = \liminf_{r\to\infty} \frac{N(r, a) - \overline{N}(r, a)}{T(r, f)}$$

$$\Theta(a, f) = 1 - \limsup_{r\to\infty} \frac{\overline{N}(r, a)}{T(r, f)}$$

By the First Fundamental Theorem of Nevanlinna it follows that $0 \leq \delta(a, f),\ \theta(a, f), \Theta(a, f) \leq 1$. If a is an omitted value, then $\delta(a, f) = 1$. Also, given $\epsilon > 0$, for sufficiently large r, we have

$$N(r, a) - \overline{N}(r, a) > (\theta(a, f) - \epsilon)T(r, f),$$
$$N(r, a) < (1 - \delta(a, f) + \epsilon)T(r, f),$$

and thus

$$\overline{N}(r, a) < (1 - \delta(a, f) - \theta(a, f) + 2\epsilon)T(r, f).$$

That is,

$$\Theta(a, f) \geq \delta(a, f) + \theta(a, f).$$

A value a is called deficient w.r.t.f if $\delta(a, f) > 0$, it is called maximally deficient if $\delta(a, f) = 1$(follows that an omitted value is maximally deficient), and it is called ramified if $\theta(a, f) > 0$. We call a value defective if either it is deficient or ramified. Now we present the Defect Relation.

Theorem 9. *The set $T = \{a : \Theta(a, f) > 0\}$ is countable, and on summing over all such values a, we have*

$$\sum_{a\in T}(\delta(a, f) + \theta(a, f)) \leq \sum_{a\in T} \Theta(a, f) \leq 2.$$

The defects measure how far the Fundamental Theorem of Algebra is from holding for f at the value a, and the Defect Relation says in some sense that the analogue of the Fundamental Theorem of Algebra fails by a finite amount, and supplies a quantative bound for the failure. The Defect Relation also proves the Picard's Little Theorem: Suppose a transcendental meromorphic function f takes three distinct values $a_1, a_2, a_3 \in \overline{\mathbb{C}}$ only finitely many times. Then $\overline{N}(r, a_k) = O(\log r)$ for $k = 1, 2, 3$, which implies that $\frac{\overline{N}(r,a_k)}{T(r,f)} \to 0$, as$r \to \infty$ which further implies that $\sum_{k=1}^{3} \Theta(a_k, f) = 3$ and this contradicts the Defect Relation.

For more rigrous treatment of defects we urge the reader to refer to Chapter 4 of Hayman's book[30], and of course Cherry and Ye[13], Lo[46] and Nev[56].

7.2 Functions Sharing Values

Two meromorphic functions f and g are said to share the value $a \in \mathbb{C}$ iff $f^{-1}(\{a\}) = g^{-1}(\{a\})$. That is, iff $\{z \in \mathbb{C} : f(z) = a\} = \{z \in \mathbb{C} : g(z) = a\}$. Perhaps some of the most striking applications of Nevanlinna value distribution theory are to the sharing of values by two meromorphic functions. The most fascinating result in this direction is the following five point theorem of Nevanlinna. For real functions, there is nothing that even remotely corresponds to this theorem.

Theorem 10. *Suppose that f and g are two meromorphic functions in the plane sharing five distinct values. Then $f \equiv g$, or f and g are both constant.*

The number five in the above theorem is sharp. For example the functions $f_1(z) = e^{-z}$, $f_2(z) = e^z$, with $a = 0, 1, -1, \infty$ show that here five cannot be replaced by four. In a special case when $g = f'$, if f and f' share two distinct finite values then $f = f'$; the number 2 in this case is also sharp: the function

$$f(z) = e^{e^z} \int_0^z e^{-e^t}(1 - e^t)dt$$

shares the value 1 with its derivative and is different from it since $\frac{f'-1}{f-1} = e^z$.

This is a quite active area of research in the Nevanlinna Theory and some of the recent accomplishments in this area are [1], [15], [34], [36], [40], [45], [69] and [76].

As a variation of this idea let f be a non-constant meromorphic function on $\mathbb{C}$. Define $E_S(f) = \bigcup_{a \in S}\{z : f(z) = a\}$, here zero of $f(z) - a$ is counted with due regard to its multiplicity. Similarly, we define $\overline{E}_S(f)$ where a zero of $f(z) - a$ is counted ignoring multiplicity. A subset S of $\mathbb{C}$ such that for any two meromorphic functions f and g such that $E_S(f) = E_S(g) \Rightarrow f \equiv g$ is called a *Unique Range Set* of meromorphic functions counting multiplicity. Similarly, the unique range set ignoring multiplicity. Currently this is very interesting and active topic of research in the Nevanlinna Theory. Following are the still unsettled problems in this area:

(A) To provide a necessary and sufficient conditions for a set to be a unique range set for meromorphic functions.
(B) To find the minimal cardinality of the Unique range set of meromorphic functions.

There is a rapid progress in this area, we refer the reader to the work of G. Frank and M. Reinders[17], Q. Han and H-X Yi([29]), P. Li and C. C. Yang([43] and [44]) and E. Mues and M. Reinders[50].

7.3 Complex Dynamics

Complex Dynamics is a thrust area in modern function theory and two consective Fields Medals in 1990s were awarded to the mathematicians for their works on Complex Dynamics. The Nevanlinna Theory also plays a vital role in the study of Dynamics of transcendental meromorphic functions through Fixed-point Theory and Normality of meromorphic functions both of which are most indispensable for the study of transcendental dynamics. Without going into the details we shall simply give some examples illustrating the applications of Nevanlinna theory to Complex Dynamics. For details on the subject we refer the reader to [3], [7], [8], [15], [35], [48], [49] and [71].

If f is holomorphic function, define a sequence $\{f^n\}$ of iterates of f as $f^1 = f$ and $f^n = fof^{n-1}$ for $n \geq 2$. The main problem in the complex dynamics is to study the behavior of the sequence $\{f^n\}$ as $n \to \infty$. A point z_0 is called a fixed-point of f if $f(z_0) = z_0$; it is called a fixed-point of exact order n of f if $f^n(z_0) = z_0$ and z_0 is not a fixed-point of any f^k for $k < n$. I. N. Baker[2] was first to use the Nevanlinna Theory to study the fixed-points of entire functions. We state here the Bergweiler's result[5] which has improved Baker's result.

Theorem 11. *If f is a transcendental entire function and n is any integer greater than or equal to 2, then f has infinitely many fixed-points of exact order n.*

Another route through which Nevanlinna Theory has found its applications in Complex Dynamics is normal families of meromorphic functions. A family $\boldsymbol{F}$ of meromorphic functions defined on a domain D is said to be *normal* in D if every sequence $\{f^n\}$ in $\boldsymbol{F}$ has a subsequence which converges uniformly on compact subsets of D to a function f which can be a constant function ∞. A subset of D on which the sequence of iterates of a meromorphic function f forms a normal family is called a *Fatou Set* of f and the compliment of Fatou set is called the *Julia Set* of f. These two subsets are the main focus of attention in complex dynamics. David Drasin[16] was first to apply Nevanlinna theory in the study of normal families of meromorphic functions. For example, he proved Hayman's conjecture[31] by using standard arguments of Nevanlinna theory. In fact he proved

Theorem 12. *Let $\boldsymbol{F}$ be a family of holomorphic functions in the unit disk Δ, and for a fixed $n \geq 3$ and $a \neq 0$ suppose that $f' - af^n = b,\ f \in \boldsymbol{F}$ has no solution in Δ. Then $\boldsymbol{F}$ is normal.*

For details on applications of Nevanlinna theory to normal families the reader is referred to Jeol Schiff's book[67], and W. Schwick's articles[65] and [66].

7.4 Factorization of Meromorphic Functions

Another branch of function theory that has come out as an application of Nevanlinna theory is the Factorization Theory of meromorphic functions. The factorization theory of meromorphic functions of one complex variable is to study how a given meromorphic function can be factorized into other simpler meromorphic functions in the sense of composition. In number theory, every natural number can be factorized as a product of prime numbers. Therefore, prime numbers serve as building blocks of natural numbers and the theory of prime numbers is one of the main subjects in number theory. In our situation, we also have the so-called prime functions which play a similar role in the factorization theory of meromorphic functions as prime numbers do in number theory. More specifically, factorization theory of meromorphic functions essentially deals with the primeness, pseudo-primeness and unique factorizability of a meromorphic function. We have the following definition.

Definition 2. *Let F be a meromorphic function. Then an expression*

$$F(z) = f(g(z)) \tag{7.1}$$

where f is meromorphic and g is entire (g may be meromorphic when f is a rational function) is called a factorization of F with f and g as its left and right factors respectively. F is said to be non-factorizable or prime if for every representation of F of the form (7.1) we have that either f is bilinear or g is linear. If every representation of F of the form (7.1) implies that f is rational or g is a polynomial(f is bilinear whenever g is transcendental, g is linear whenever f is transcendental), we say that F is pseudo-prime (left-prime, right-prime respectively). If the factors are restricted to entire functions, the factorization is said to be in entire sense and we have the correspoding concepts of primeness in entire sense(called E-primeness), pseudo-primeness in entire sense (called E-pseudo-primeness) etc.

According to the Definition 2, it seems inevitable that one must consider meromorphic factors in a factorization in order to determine whether a given entire function is prime or not. However, F. Gross(see [15] Lemma 3.1, pp. 116) has proved that any non-periodic entire function is prime if and only if it is E-prime. Thus for a certain class of entire functions we need not to consider meromorphic factors. Rosenbloom[62] was first to introduce the concept of prime entire functions while investigating the fixed-points of iterates of transcendental entire functions. He stated without proof that the function $z+e^z$ is prime and remarked that the proof was complecated. In 1968 Fred Gross (see [18] or [19]) gave a complete definition of factorization of meromorphic function. He not only proved the primeness of $z+e^z$ but started a series of studies on factorization of meromorphic functions (see [18] to [22], [23] to [25] and [26]). Also, he raised a very famous conjecture[18] known as Gross's Conjecture: $z + P(z)e^{\alpha(z)}$ *is prime, where $P(z)$ is a polynomial and $\alpha(z)$ is an entire function.* The fixed-point version of Gross's Conjecture is: *Let f and g be two non-linear entire functions, at least one of them being transcendental. Then the composite function fog has infinitely many fixed-points.* This conjecture remained unsettled for more than two decades and was proved in affirmation by Bergweiler[4] in 1990. Though there is abundance of research/open problems in the theory but the following conjecture of He and Yang[32] is the most challenging one.

He-Yang Conjecture: *Let f be a pseudo-prime transcendental meromorphic function and P be a polynomial of degree at least* 2. *Then foP is pseudo-prime.*

For deep insight into this area the reader is referred to the books by Chuang and Yang[15] and Gross[20]. For further developments in this area one may refer to [6], [9], [10], [11], [57], [58], [61], [68], [72], and [77]. For Factorization theory of meromorphic functions of several variables one may refer to [42] and [78].

7.5 Complex Differential Equations

Nevanlinna theory has been applied to get insight into the properties of solutions of complex differential equations. The first such applications were made by F. Nevanlinna[51] to study the meromorphic solutions with maximum deficiency sum of the differential equation $f'' + A(z)f = 0$ with $A(z)$ a polynomial. Also R. Nevanlinna[55] considered the same differential equation in connection of covering surfaces with finitely many branch points. K. Yosida[80] in 1933 proved the celebrated Malquist theorem[47] by using Nevanlinna theory. Malquist theorem states that a differential equation of the form $y' = R(z, y)$, where the right-hand side is rational in both arguments, which admits a transcendental solution, reduces to a Riccati differential equation

$$y' = a_0(z) + a_1(z)y + a_2(z)y^2$$

with rational coefficients. Since 1933 a whole range of theorems in complex differential equations can be classified as "Malquist-Yosida" theorems. For getting deep insight into the subject we refer the reader to Ilpo Laine's book[37].

We conclude our discussion on Nevanlinna Theory by mentioning that recently R. G. Halburd and R. J. Korhonen (see [27] and [28]) have extended Nevanlinna Theory to a theory for the exact difference $f \mapsto \Delta f = f(z+c) - f(z)$, where f is a meromorphic function and c is a fixed constant.

Acknowledgement

I am thankful to Prof. A. P. Singh who read the manuscript and made many valuable comments towards the improvement of the article.

References

[1] T. C. Alzahary, *Small functions and weighted sharing three values,* Complex Variables, **50**, (2005), 1105–1124.

[2] I. N. Baker, *The existence of fix-points of entire functions,* Math. Z. **73**, (1960), 280–284.

[3] A. F. Beardon, *Iteration of rational functions,* Springer, New York, Berlin, Heidelberg, 1991.

[4] W. Bergweiler, *Proof of a conjecture of Gross concerning fix-points,* Math. Z. **204**, (1990), 381–390.

[5] W. Bergweiler, *Periodic points of entire functions: proof of a conjecture of Baker,* Complex Variables Theory Appl. **17**, (1991), 57–72.

[6] W. Bergweiler, *On factorization of certain entire functions,* J. Math. Anal. Appl. **193**, (1995), 1003–1007.

[7] W. Bergweiler, *Iterations of meromorphic functions,* Bull. Amer. Math. Soc.(N.S.) **29**, (1993), 151–188.

[8] L. Carleson and T. W. Gamelin, *Complec dynamics,* Springer, New York, Berlin, Heidelberg, 1993.

[9] K. S. Charak and I. Laine, *On a class of prime entire functions,* Complex Variables, **43**, (2001), 251–258.

[10] K. S. Charak and I. Laine, *On a class of prime entire functions,* To appear in Acta Math. Sinica.

[11] K. S. Charak, I. Laine and D. Rochon, *Factorization of bicomplex meromorphic functions,* To appear in Trends in Maths. Birkhauser(Basel), 2008.

[12] Z. Chen and Q. Yan, *Uniqueness problem of meromorphic functions sharing small functions,* Proc. Amer. Math. Soc. **134**, (2006), 2895–2904.

[13] W. Cherry and Z. Ye, *Nevanlinna's theory of value distribution,* Springer, New York, Berlin, Heidelberg, 2001.

[14] C. T. Chuang, *une generalisation d'une inegalite de Nevanlinna,* Sci. Sinica, **13**, (1964), 887–895.

[15] C. T. Chuang and C. C. Yang, *Fix-points and factorization of meromorphic functions,* World Scientific Publishing Co. Pvt. Ltd. Singapore, 1990.

[16] D. Drasin, *Normal families and the Nevanlinna theory,* Acta Math. **122**, (1969), 231–263.

[17] G. Frank and M. Reinders, *A unique range set for meromorphic functions with* 11 *elements,* Complex Variables Theory Appl. **37** (1998), 185–193.

[18] F. Gross, *On factorization of meromorphic functions,* Lectures at the Math. Instt. University of California La Jolla, June 28, 1966.

[19] F. Gross, *On factorization of meromorphic functions,* Trans. Amer. Math. Soc. **131**, (1968), 215–222.

[20] F. Gross, *Factorization of meromorphic functions,* Math. Res. Center, Naval Res. Lab. Washington D.C. 1972.

[21] F. Gross, *Prime entire functions,* Trans. Amer. Math. Soc. **161**, (1971), 219–233.

[22] F. Gross, *Factorization of meromorphic functions and some open problems,* Lecture Notes, **599** Springer-Verlag, New York, 1976.

[23] F. Gross and C. C. Yang, *The fix-points and factorization of meromorphic functions,* Trans. Amer. Math. Soc. **168**, (1972), 211–219.

[24] F. Gross and C. C. Yang, *Further results on prime entire functions,* Trans. Amer. Math. Soc. **142**, (1974), 347–355.

[25] F. Gross and C. C. Yang, *On pseudo-prime entire functions,* Tohoku Math. J. **26**, (1974), 65–71.

[26] F.Gross, C.C. Yang and C.F. Osgood, *Primable entire functions,* Nagoya Math. J. **51**, (1973), 123–130.

[27] R. G. Halburd and R. J. Korhonen, *Difference analogue of the lemma on the logarithmic derivative with applications to difference equations,* To appear in J. Math. Anal. Appl. 2005, arxiv: math/0504245v1.

[28] R. G. Halburd and R. J. Korhonen, *Nevanlinna theory for the difference operator,* arxiv: math/0506011v1.

[29] Q. Han and H-X Yi, *Some further results on meromorphic functions that share two sets,* Ann. Polon. Math. **93**, (2008), 17–31.

[30] W. K. Hayman, *Meromorphic functions,* Clarendon Press, Oxford, 1964.

[31] W. K. Hayman, *Research problems in function theory,* London, 1967.

[32] Y. Z. He and C. C. Yang, *On pseudo-primality of the product of some pseudo-prime meromorphic functions,* Analysis of one complex variable, edited by C.C. Yang, World Scientific Publishing Co. Pvt. Ltd. Singapore, (1987), 113–124.

[33] A. Hinkkanen, *A sharp form of Nevanlinna's second fundamental theorem,* Invent. Math. **108**, (1992), 549–574.

[34] P. C. Hu, P. Li and C. C. Yang, *Unicity of meromorphic functions,* Kluwer Acad. Publ., Dordrecht-Boston-London, 2003.

[35] P. C. Hua and C. C. Yang, *Dynamics of transcendental functions,* Gordon and Breach Science Publishers, 1998.

[36] A. H. H. A. Khaladi, *On entire functions which share one small function CM with their derivative,* Result. Math. **47**, (2005), 1–5.

[37] I. Laine, *Nevanlinna theory and differential equations,* Walter de Gruyter, Berlin, New York, 1993.

[38] S. Lang, *The error term in Nevanlinna theory,* Duke Math. J. **56**, (1988), 193–218.

[39] S. Lang and W. Cherry, *Topics in Nevanlinna theory III,* Lect. Notes Math. **1433** Springer, New York (1990).

[40] I. Lahiri and N. Mandal, *Meromorphic functions sharing a single value with unit weight,* Kodai Math. J. **29**, (2006), 41–50.

[41] P. Li, *Entire functions that share two values,* Kodai Math. J. **28**, (2005), 551–558.

[42] B. Q. Li and C. C. Yang, *Factorization of meromorphic functions in several complex variables,* AMS Contemp. Math. **142**, (1993), 61–73.

[43] P. Li and C. C. Yang, *Some further results on the unique range sets of meromorphic functions,* Kodai Math. J. **18**, (1995), 437–450.

[44] P. Li and C. C. Yang, *On the unique range set of meromorphic functions,* Proc. Amer. Math. Soc. **124**, (1996), 177–185.

[45] S. Lin and W. Lin, *Uniqueness of meromorphic functions concerning weakly weighted-sharing,* Kodai Math. J. **29**, (2006), 269–280.

[46] Y. Lo, *Value distribution theory,* Springer-Verlag, Berlin, Heidelberg, New York and Science Press, Beijing, 1993.

[47] J. Malquist, *Sur les fonctions a'un nombre fini des branches definies par les equations differentielles du premier ordre,* Acta Math. **36**, (1913), 297–343.

[48] J. Milnor, *Dynamics in one complex variable: introductory lectures,* Vieweg, 2000.

[49] S. Morosawa, Y. Nishimura, M. Taniguchi, and T. Ueda, *Holomorphic dynamics,* Cambridge University Press, Cambridge, 2000.

[50] E. Mues and M. Reinders, *Meromorphic functions sharing one value and unique range sets,* Kodai Math. J. **18**, (1995), 515–522.

[51] F. Nevanlinna, *Uber eine klasse meromorpher funktionen,* Septieme congres Math. Scand. Oslo, (1929), 81–83.

[52] R. Nevanlinna, *Sur les valeurs exceptionnelles des fonctions meromorphes,* C.R. Academy, Paris, **179**, (1924), 24–26.

[53] R. Nevanlinna, *Zur theorie der merophen funktionen,* Acta Math. **45**, (1925), 1–99.

[54] R. Nevanlinna, *Le theoreme de Picard-Borel et la theorie des fonctions meromorphes,* Paris, 1929.

[55] R. Nevanlinna, *Uber Riemannschen Flachen mit endlich veilen Windungspunkten,* Acta Math. **58**, (1932), 295–373.

[56] R. Nevanlinna, *Analytic functions,* Springer-Verlag, New York, Berlin, Heidelberg, 1970.

[57] T. W. Ng and C. C. Yang, *On the composition of a prime transcendental function and a prime polynomial,* Pacific J. Math. **193** (2000), 131–141.

[58] Y. Noda, *On factorization of entire functions,* kodai Math. J. **4**, (1981), 480–494.

[59] O. Lehto, *On the birth of the Nevanlinna theory,* Ann. Acad. Scien. Fenn. **7** (1982), 5–23.

[60] C. F. Osgood, *Sometimes effective thue-Siegel-Roth-Schmidt-Nevanlinna bounds, or better,* J. Number Theory, **21**, (1985), 347–389.

[61] M. Ozawa, and K. Sawada, *A theorem on factorization of certain entire functions,* Complex Variables, **34** (1997), 181–185.

[62] P. C. Rosenbloom, *The fix-points of entire functions,* Medd. Lunds Univ. Math. Sem. M. Riesz (1952), 187–192.

[63] M. Ru, *Nevanlinna theory and its relation to diophantine approximation,* World Scientific, Singapore, London, Hong Kong, 2001.

[64] L. A. Rubel, *Entire and meromorphic functions,* Springer, New York, Berlin, Heidelberg, 1996.

[65] W. Schewick, *Normality criteria for families of meromorphic functions,* J. Analyse Math. **52**, (1989), 241–289.

[66] W. Schewick, *Sharing values and normality,* Arch. Math. **59**, (1992), 50–54.

[67] J. L. Schiff, *Normal families,* Springer-Verlag, New York, Berlin, Heidelberg, 1993.

[68] Anand P. Singh and K. S. Charak, *Factorization of certain classes of entire and meromorphic functions,* Indian J. pure appl. Math. **28**(1), (1997), 81–93.

[69] M. Shirosaki, *On meromorphic functions sharing three values and a set,* Kodai Math. J. **29**, (2006), 475–484.

[70] N. Steinmetz, *Uber eine Verallgemeinerung des zweiten Nevanlinnasche Hauptsatzes,* J. Reine Angew. Math. **368**, (1986), 134–141.

[71] N. Steinmetz, *Rational iteration,* Walter de Gruyter, Berlin, 1993.

[72] H. urabe, *Some further results on factorization of entire functions,* Complex Variables, **43** (2001), 457–467.

[73] P. Vojta, *Diophantine approximations and value distribution theory,* Lect. Notes Math. Springer, New York, **1239**, 1987.

[74] H. Weyl and F. J. Weyl, *Meromorphic functions and analytic curves,* Princeton Univ. Press, Princeton, 1943.

[75] P. Wong, *On the second main theorem in Nevanlinna theory,* Amer. J. Math. **111**, (1989), 549–583.

[76] L. Z. Yang, *Meromorphic functions and also their first two derivatives have the same zeros,* Ann. Acad. Sci. Fenn. Math. **30**, (2005), 205–2118.

[77] C. C. Yang, *Some aspects of factorization theory: A survey,* Complex Variables, **13**, (1989), 133–142.

[78] C. C. Yang, *Factorization theory of meromorphic functions of one and several complex variables,* Complex Analysis and its applications, Longman Scientific and Technical, (1994), 114–128.

[79] Z. Ye, *On Nevanlinna's error terms,* Duke Math. **64**, (1991), 243–260.

[80] K. Yosida, *A generalization of Malquist's theorem,* J. Math. **9**, (1933), 253–256.

[81] G. H. Zhang, *Theory of entire and meromorphic functions,* AMS, 1993.

Truth vs. Provability in Mathematics: The Completeness Theorem

Ayalur Krishnan

Department of Mathematics, Kingsborough Community College, Cuny

E-mail: akrishnan@kingsborough.edu

Abstract. Most mathematicians study issues concerning truth and provability. For example, when they talk about number theory or analysis, they have specific structures in mind, such as the usual set of natural numbers $\mathbb{N}$, or the set of real numbers $\mathbb{R}$. And when they prove statements regarding these structures, they use particular properties of the natural or real numbers, combined with logical steps in reasoning that most people would agree are valid. Thus, there are two main components in mathematical reasoning: the formal logical part and the part that derives meaning because we are talking about specific structures such as $\mathbb{N}$ or $\mathbb{R}$. Gödel's Completeness Theorem elucidates the subtle connection between these two aspects in mathematics. We give a self-contained statement of this theorem, and in the process also give an introduction to mathematical logic.

Introduction

Mathematicians want to know the answers to particular questions regarding particular mathematical objects. For example, once we have defined the concept of "prime", we can ask whether it is true that there are infinitely many primes. And we can exhibit an elementary proof of this statement. A more complicated question would be to ask whether Fermat's Last Theorem is true. As is well known, the answer is provided by Wiles' proof. This proof is by no means easy to understand, but it has been looked at by many experts in number theory, and nobody really doubts its correctness. Of course, many people still aim to find an "elementary proof" of the theorem. A very famous question relates to the cardinality of the set of real numbers: Is there a set whose cardinality is strictly larger than that of the set of natural numbers, and strictly smaller than that of the set of real numbers? This question is known as the Continuum Hypothesis. The hypothesis is that there is no such set. It has been shown, using strictly *mathematical methods*, that this question *cannot* be answered either in the affirmative or in the negative by using the currently accepted ideas of mathematics. This is an intriguing example of a situation in mathematics where it is possible to *prove* that one cannot either *prove* or *disprove* a statement about certain mathematical objects. These issues raise the questions: What is truth, and what is provability? What are the connections between truth and provability? Is it true that all true mathematical statements can be proven? What is a proof? Are there any limitations to what can be proven?

Mathematical logic is the subject that studies these questions. The distinctive feature about mathematical logic is that *mathematical methods* are used in the analysis of questions about mathematics. Thus, mathematical logic is also called metamathematics, since it is the study of methods used in the study of mathematics. Mathematical logic is a relatively new subject. It really got off the ground only about 150 years ago. Though some of the basic principles that underlie mathematical logic are already to be found in Euclid's Elements, they were not explicitly studied as an independent subject till the 19th century. In the previous century, mathematical logic saw many stunning successes. Many mathematical questions were answered using the techniques developed in mathematical logic, and many subtle issues about the nature of mathematics itself were clarified. Mathematical logic makes it possible to give one foundational treatment for various disciplines of mathematics, such as number theory, analysis, algebra etc. In this article we will give examples of the kinds of questions that logicians study, and the methods that they use.

1. The Basic Issues Studied in Mathematical Logic

Mathematical logic tries to understand the nature of mathematical activity. We will illustrate some of the basic ideas using concepts from number theory. To start with, if we ask what the numbers are, then there will be many different answers, depending on the person asked. But rather than ask this question, we assume that the concept of "number" is an undefined, primitive, and intuitively clear concept. What we are more interested in are the *properties* of numbers, and what we can say about questions asked about numbers. We begin by giving names for certain primitive concepts. So we denote by "0" the number zero. Observe that we are not *defining* the number zero. We agree that most people have an understanding of what the concept of zero is, and we are just giving this concept a name. Similarly, we give a name "+" to the concept of addition. We can also give a name to the "successor" operation, which when applied to any number gives us the next number. This we call as "S". Once we have these names for concepts, nobody would seriously argue about the following property of numbers: $x + S(y) = S(x + y)$. So what happened here is that we gave some names for concepts, and we wrote down certain properties about these concepts that we would like to consider as true about these concepts. Such statements that we want to hold as primitively true, and not requiring any explanation, are called *axioms*. We can similarly write down certain other axioms about natural numbers. This is similar to what Euclid did: He gave a name for certain primitive concepts such as "point", "straight line" etc., and wrote down some basic axioms about these concepts. Thus, two points determining a unique straight line is an axiom in plane geometry.

Once we have written down basic axioms about the objects that we want to study, we ask what else can be gleaned from these axioms. In the example of Euclid, starting from the basic axioms about geometric objects, he was able to conclude a great many other properties about these geometric objects. How did he do so? Or in the example of numbers, we start with some basic assumptions about numbers, and then we can conclude a great many other facts about numbers. How do we go about doing this? Well, we use the concept of a proof. A proof of a statement about numbers is an argument that uses other known (proven) facts about numbers, the axioms about numbers, and *logical reasoning* to conclude new facts. So what is this concept of logical reasoning? Certain facts are true by virtue of *logic*. Suppose I say "If it is raining, the road is wet". This statement is normally considered to be true. From this, we can logically conclude: "If the road is not wet, it has not been raining". But we cannot logically conclude "If it is not raining, the road is not wet" (Why?). These kinds of inferences, which depend on the structure of statements, and not their meaning, are called logical inferences. Mathematicians use these kinds of reasoning all the time, without explicitly recognizing such a step in their argument. For example, if A implies B, and if B implies C, where A, B, and C may mean anything whatsoever, then it follows logically that A implies C. A chain of reasoning, that starts from axioms about concepts, and uses other known statements about these concepts, and logical inferences, is called a *proof*. A more formal definition will be given later.

In order to talk about concepts, we introduce the notion of a formal language. This is similar to the notion of a natural language such as English, except that the rules governing a formal language are a lot stricter. The formal language would have notation for non-logical concepts, such as 0, $+$, S, etc., and notation for logical concepts such as implications ($\Rightarrow$), negation ($\neg$) etc. There will be rules for saying what is allowed and what is not allowed to be written as a formal expression. We can now write formally the axioms, the rules of inference etc. in this formal language.

Once we have these basic ideas, we can ask further questions. For example, once we have written down the concepts about natural numbers and the axioms that are true of them, we can ask whether there is a *structure* for the formal language that *satisfies* these axioms. That is, we seek to give *meaning* to the notation that we have introduced. Informally, we seek a set, whose elements we will call natural numbers. Note that we do not care about the nature of these objects. For the purpose of understanding, we can think of a number, say seven, as a collection of seven matchsticks tied together by a rubber band, and the set of natural numbers as the set consisting of all these collections of matchsticks. Of course, there are not enough matchsticks or rubber bands to populate this set, but that is not our concern. Next we seek an interpretation of the basic concepts such as 0, S, $+$ etc. in relation to this set. Thus, we have the usual set $\mathbb{N}$ of natural numbers, and we have the usual interpretation of 0, S, $+$ etc. We satisfy ourselves that the axioms of natural numbers are true when interpreted with respect to this set. But then we can ask: are there other structures, which do not look like $\mathbb{N}$, but with respect to which also the axioms are satisfied? What does it mean to say that these other structures "do not look like" $\mathbb{N}$? What if there was a statement expressed in the formal language for natural numbers such that it would turn out to be true when interpreted in every structure for this language with respect to which the axioms were true? Would this statement be derivable from the axioms? In other words, would there be a proof of this statement? Could there be statements known to be true in this structure that cannot be proven? How then are these statements known to be true?

We now state some of the basic results in logic informally. Suppose we have a formal language for the natural numbers. And suppose we have a structure for the language with respect to which the axioms about numbers are true. Note that this structure does not have to be $\mathbb{N}$. And suppose we have a proof of a statement, say A, derived from the axioms for numbers. Then A, interpreted in this structure, will be true. This is called the Soundness Theorem. The idea of this result is that by starting from statements that are true, by using the principles of logic, one only passes to true statements, never to false statements. The converse of this result is also true, though much harder to see. It is called the Completeness Theorem. The converse is the statement that if there is a statement A, say in the language of natural numbers, that is true in every structure for the language of natural numbers with respect to which the axioms about natural numbers are true, then this statement has a proof from the axioms of natural numbers. Note that this statement is not saying that every true statement about the natural numbers (with respect to the structure $\mathbb{N}$) can be proven using the axioms of natural numbers.

One other important result in logic is the Incompleteness Theorem. The use of the word "complete" in this and the Completeness Theorem are not technically related. The Incompleteness Theorem states roughly that it is not possible to have an algorithm that lists all the true facts about the numbers in $\mathbb{N}$. In other words, if we came up with any algorithm no matter what, that starts to list the true statements about natural numbers, then it is possible to obtain one true statement about the numbers that will not be contained in this list. This is basically a very clever diagonalization argument. So this result essentially says that computers cannot replace human ingenuity and creativity, though they can be very fast. These results, though often misunderstood as saying that there are true statements about natural numbers that can *never* be proven, do point out serious limitations in the way we view mathematical proof and truth, and have sweeping philosophical and mathematical consequences. Whether there are any methods to overcome this limitation is a fascinating question.

Finally, we state the notion of independence in relation to the Continuum Hypothesis that we stated earlier. We fix a formal language for sets, and we figure out how to talk about the basic notions of sets, such as subsets, power sets, infinite sets etc. We can also write down axioms about sets in the formal language. Note now that it is tricky to talk about a structure for the axioms. Here's why. A structure for a mathematical theory is a set. So a structure for the theory of sets should be a set with respect to which we can interpret the axioms of sets. Also, one would imagine that all relevant sets that we want to discuss live inside this set. But then how do we interpret the power set of this set, clearly a relevant set? These issues are not fully resolved yet, but it is possible to talk about structures for set theory in a more technical way where we can satisfactorily deal with issues such as the one above. Once we have this machinery, we can talk about sets such as $\mathbb{N}$, $\mathbb{R}$ etc. And we can phrase questions about cardinality of these sets. Thus, we can formally state the Continuum Hypothesis, and ask whether it can be proven using the axioms of set theory. The surprising answer is that one can exhibit a structure for set theory where the Continuum Hypothesis is true and a structure where it is false. Thus, by the Soundness Theorem, it is not possible that the Continuum Hypothesis has a proof from the axioms of set theory (Why?). We say that the Continuum Hypothesis is *independent* of the axioms of set theory.

And now we name a couple of names. The Soundness and Completeness Theorems, as well as the Incompleteness Theorem were proven by Kurt Gödel. He proved the Completeness Theorem in his PhD thesis in 1929 at the age of 23, and the Incompleteness Theorem in 1931 at the age of 25. Gödel is one of the most fascinating figures in the history of ideas. He also proved one half of the independence of the Continuum Hypothesis: he proved that it is possible to exhibit a structure for set theory with respect to which the hypothesis is true. The other half of the independence, that there is a structure for set theory with respect to which the hypothesis is false, was proven by Paul Cohen. Cohen got a Fields Medal for his work in 1966, the only Fields Medal ever awarded for a work in logic.

2. The Formal Setup

We will use basic ideas from number theory to motivate and illustrate our concepts. We will follow the treatment as in Shoenfield's book [3]. Since this is an informal account we will attempt to make the concepts as clear as possible without too much symbolism, so that we don't lose our reader. If the reader's interest is piqued after reading this article, we strongly urge the reader to read Shoenfield's classic for a deep treatment of the subject.

Definition 2.1. *A* first order language *consists of three parts:*

(i) *An infinite list of variables,* $x, y, z, x', y', z', x'', y'', z'', \ldots$.
(ii) *For each n, n-ary function symbols, and n-ary predicate symbols. Among the predicate symbols, we assume that the* 2*-ary predicate "$=$" always exists.*
(iii) *The symbols* $\neg$, $\vee$, *and* $\exists$.

The symbols in (i) and (iii) are called *logical* symbols. The symbols in (ii) are called the *non-logical* symbols. The logical symbols in (iii) stand, respectively, for "not", "or" and "there exists". You may have seen other logical symbols, such as $\forall$ (universal quantifier), $\Rightarrow$ (implication), $\Leftrightarrow$ (iff) etc. These can all be expressed using the three we have given. For example, if it is true that all people in Bangalore love logic, then it is also true that it is not the case that there exists someone in Bangalore who does not love logic. In other words, $\forall x A$ ("for all x A is true") is the same as $\neg\exists x \neg A$ (it is not the case that there exists an x for which A is not true). The symbol for "and" ($\wedge$) can be obtained from that for "or" by negation and De-Morgan's laws. The symbols we have given constitute the most economic possible set.

The non-logical part consists of symbols specific to the topic under discussion. Thus, if we are discussing the natural numbers, we will have the symbols 0, S, $+$, $*$, and $<$, to stand for the number zero, the successor function, addition, multiplication, and the relation "less than", respectively. The functions symbols and predicate symbols are used to denote these kinds of concepts. For example, the successor function symbol S is a 1-ary function symbol, $+$ a 2-ary function symbol, and $<$ a 2-ary predicate symbol. We can use "$\in$" to denote the membership predicate for sets when we speak about set theory. This is a 2-ary predicate symbol. The logical symbols are the same for all topics under discussion, be they groups, or numbers, or Hilbert spaces, whereas the non-logical symbols are specific to the subject under discussion

The phrase "first-order" in definition 2.1 is used to indicate that the variables range over the particular objects under investigation, but not over subsets thereof. Thus, while speaking about numbers, the variables are supposed to range over numbers, and not over subsets of numbers. In order to talk about the subset of prime numbers, we have to first interpret number theory in set theory. Observe that in set theory, since the variable are supposed to range over sets, there is no problem with talking about subsets etc. This point is connected with the issue that we are using set theory to study logic and making set theory itself an object of study in logic. We refer the reader to [3] and [1] for more details about this delicate issue.

We use the logical and non-logical symbols to construct terms and formulas. Thus, $S(S(0))$ is a term in the language for natural numbers, and $\exists x(y = S(S(S(x))))$ is a formula of the language. As you can probably tell, the terms are intended to represent particular objects under investigation, whereas formulas are intended to convey assertions about these objects.

Definition 2.2. *A* term *is given by the inductive definition:*

(i) *A variable is a term.*
(ii) *If* $u_1, \ldots, u_n$ *are terms, and* f *is an n-ary function symbol, then* $f u_1 \ldots u_n$ *is a term.*

Definition 2.3. *A* formula *is given by the inductive definition:*

(i) *An atomic formula is of the form* $pu_1 \ldots u_n$ *where* p *is an n-ary predicate symbol, and* $u_1, \ldots, u_n$ *are terms.*
(ii) *If* A *is a formula, so is* $\neg A$.
(iii) *If* A *and* B *are formulas, so is* $\vee AB$.
(iv) *If* A *is a formula, so is* $\exists x A$ *where* x *is a variable.*

Note that we use parentheses and other notational conventions only to make reading easier, but these are not part of the definition of the formal language that we have given above. Thus, the correct way of writing $x + y$ would be, as per (ii) of 2.2, "$+xy$" but this is hard on the eye, so we use the infix notation and write $x + y$, remembering that this is really a notational convention.

Definition 2.4. *Let L be a first order language. A* structure *for L consists of:*

(i) *A non-empty set* X.
(ii) *For each n-ary function symbol* f *of L, an n-ary function* f_X *from* X *to* X.
(iii) *For each n-ary predicate symbol* p *of L, an n-ary relation* p_X *on* X.

We illustrate the previous definitions using the case of natural numbers. The formal first-order language for speaking about natural numbers is given by the non-logical symbols: 0 (which is a constant, and thus can be viewed as a 0-ary function symbol), the 1-ary function symbol S, the 2-ary function symbols $+$ and $*$, and the 2-ary predicate symbol $<$. These symbols have the evident meaning associated with them. Note that we only need to specify the non-logical symbols, since the logical symbols are common for every situation. A structure for this language would be $\mathbb{N}$, where the non-logical symbols have the usual interpretations. The symbol "0" is interpreted as the number 0, the successor function S is interpreted as a function from $\mathbb{N}$ to $\mathbb{N}$ that does the expected thing ($Sx = x + 1$; of course, we do not have "1" in our language, so the correct formal way to say what S does would be to say that $Sx = x + S0$), and $<$ is interpreted in the usual way as a relation on $\mathbb{N} \times \mathbb{N}$. Also try and convince yourself that we do really need all the symbols we have introduced. For example, we cannot do without "S": try and express what S does by just using the other symbols.

We next introduce the concept of free and bound occurrences of variables. Suppose we write $\exists x \forall y (y < x)$. This is intended to be read as: "There exists an x such that for all y it is the case that y is less than x". For this statement to be true in the structure $\mathbb{N}$, there would have to exist one number which is greater than all numbers. This statement is obviously false when referring to the structure $\mathbb{N}$. The occurrences of the variable x are said to be "bound" in the formula $\exists x \forall y (y < x)$ because there is a quantifier having the variable x immediately following it, and because this quantifier affects the second occurrence of x also. But in the formula $\exists y (y < x)$, we do not specify any restrictions on x. Thus it is "free". In this formula, y is bound. For this formula to be true with respect to $\mathbb{N}$, the meaning of the formula would have to be true no matter which natural number we think of x as representing. $\exists y (y < x)$ is false when applied to $\mathbb{N}$. Can you tell why? It is possible for one occurrence of a variable to be free and for the other to be bound in the same formula. Here is an example: $(y < x) \vee (\exists x \forall y (y < x))$. Here the first occurrence of x is free, and the other occurrences are bound. Once we have the notion of free and bound occurrences, we denote by $A_x[u]$ the result of replacing every free occurrence of the variable x in the formula A by the term u.

With the preceding ideas, here is a list of (non-logical) axioms (formulas) about natural numbers written in the formal language. The variables in the list are supposed to range over natural numbers. We repeat that we use parentheses and logical symbols other than $\neg$, $\forall$, and $\exists$ only to avoid confusion and that they are not part of the formal language.

N_1 $\quad \neg(Sx = 0)$
N_2 $\quad Sx = Sy \Rightarrow x = y.$
N_3 $\quad x + 0 = x.$
N_4 $\quad x + Sy = S(x + y).$
N_5 $\quad x * 0 = 0.$
N_6 $\quad x * Sy = (x * y) + x.$
N_7 $\quad \neg(x < 0).$
N_8 $\quad x < Sy \Leftrightarrow x < y \vee x = y.$
N_9 $\quad x < y \vee x = y \vee y < x.$
N_{10} $\quad (A_x[0] \wedge \forall x (A \Rightarrow A_x[Sx])) \Rightarrow A.$

Definition 2.5. *A formula is said to be* valid *in a structure if its meaning is true when interpreted in that structure.*

At this point it is a good exercise for the reader to go through the list of the axioms $N_1 - N_{10}$ and convince herself that these axioms are valid when interpreted in the structure $\mathbb{N}$. Observe that N_{10} is a formulation of the induction axiom, and is actually a schema, one formula for each A. In this sense, the list $N_1 - N_{10}$ is an infinite list, because there is a different statement of N_{10} for each A.

We now present the logical axioms. Remember that these axioms are valid in all structures. In fact, you should convince yourself that these axioms will indeed be valid in all structures because of their logical nature. The letter A ranges over formulas, and "f" and "p" in L_4 and L_5 range over function symbols and predicate symbols, respectively. Also remember that some of the symbols that we use are not technically part of the language, but can be written in terms of the ones that are part of the formal definition of a language. We use these abbreviations for the sake of readability.

L_1 $\quad \neg A \vee A.$
L_2 $\quad A_x[a] \Rightarrow \exists x A.$
L_3 $\quad x = x.$
L_4 $\quad x_1 = y_1 \wedge \ldots \wedge x_n = y_n \Rightarrow f x_1 \ldots x_n = f y_1 \ldots y_n.$
L_5 $\quad x_1 = y_1 \wedge \ldots \wedge x_n = y_n \Rightarrow p x_1 \ldots x_n \Rightarrow p y_1 \ldots y_n.$

And we finally present the logical rules of inference. It is important to note that if the hypotheses of any of these inference rules are valid in some structure, then the conclusion of that rule is also valid in that structure. This should of course be the case, otherwise how could we use logical inference? It is straightforward in the first four rules to see that this is the case, but the fifth rule requires some thought.

LR_1 Infer $B \vee A$ from A.
LR_2 Infer A from $A \vee A$.
LR_3 Infer $(A \vee B) \vee C$ from $A \vee (B \vee C)$.
LR_4 Infer $B \vee C$ from $A \vee B$ and $\neg A \vee C$.
LR_5 If x is not free in B, infer $\exists x A \Rightarrow B$ from $A \Rightarrow B$.

Definition 2.6. *A* first-order theory *or simply a* theory *is a formal system consisting of:*

(i) *A first-order language.*
(ii) *The logical and non-logical axioms.*
(iii) *The logical rules of inference.*

We can now speak of the first-order theory of natural numbers. This will have the formal language that we introduced earlier for speaking about natural numbers, the non-logical axioms N_1 through N_{10}, the logical axioms L_1 through L_5, and the logical rules of inference LR_1 through LR_5.

Definition 2.7. *A* model *of a theory is a structure with respect to which the non-logical axioms of that theory are valid.*

One question is whether there are any other models for the axioms $N_1 - N_{10}$. The (perhaps surprising) answer is that there are many other models, even uncountable models. In fact, there are models of the axioms $N_1 - N_{10}$ where "infinite" elements exist: i.e. elements x such that x is simultaneously greater than $S0$, $SS0$, $SSS0$, and so on. This leads to a whole active current topic of research in logic called non-standard number theory.

3. Soundness and Completeness Theorems

We can now state the soundness and completeness theorems. We first need to introduce the formal concept of a proof and that of a theorem.

Definition 3.1. *A (formal)* proof *in a theory is a finite list of formulas, each formula of the list satisfying:*

(i) *It is either a logical axiom or a non-logical axiom of the theory.*
(ii) *It is the conclusion of a logical inference rule applied to hypotheses that appear before this formula on the list.*

Definition 3.2. *A (formal)* theorem *of a theory is the last formula of a proof in that theory.*

Definition 3.3. *A formula is said to be valid* in a theory T *if it is valid in every model of T.*

We can now state the Soundness Theorem.

Theorem 3.1 (Soundness). *If T is a theory, then every theorem of T is valid in T.*

Proof. The proof is by induction on theorems. A non-logical axiom of the theory is valid in any model of the theory by the definition of a model. A logical axiom is valid in all structures. If the theorem is the conclusion of an inference rule, then the result is true by the induction axiom the comments immediately preceding the list of inference rules. □

The idea of this result is straightforward: Imagine we have the theory of natural numbers given by the non-logical axioms $N_1 - N_{10}$. Now consider any model of these axioms, such as the structure $\mathbb{N}$. Then, the soundness theorem says that if we were to generate more theorems from $N_1 - N_{10}$, along with the help of the logical axioms and the logical rules of inference, we would arrive at formulas that remain true in $\mathbb{N}$. Surely we would want a result like this, otherwise we would not have a guarantee that the act of doing everyday mathematics does not lead to wrong results.

Theorem 3.2 (Completeness). *If a formula of a theory T is valid in T, then it is a theorem of T.*

The statement of this result is easy to understand: imagine that there is a formula A in the language of number theory that we introduced earlier, involving S, 0, $<$, $+$, $*$, and the logical symbols. Now imagine that the formula A is true when interpreted in every model of the axioms $N_1 - N_{10}$, not just in $\mathbb{N}$. Then the completeness theorem says that it is possible to have a proof of this formula A, that is, a finite list consisting of formulas of the form $N_1 - N_{10}$, and using the logical axioms and the logical rules of inference. This is far from obvious.

Gödel also proved the Compactness theorem, which can be shown to be equivalent to the Completeness theorem. This requires some advanced ideas. But the statement of the Compactness Theorem is easy to understand, and it has startling consequences. The Compactness Theorem says that if there is an infinite set **S** of formulas such that for every finite subset **S**′ of **S**, there exists a structure relative to which every formula in **S**′ is valid, then there is a structure relative to which all of the formulas in **S** are valid. To see some interesting consequences, suppose the set **S** consists of the formulas $\exists x(x > S^n(0))$, one for each "n", where S^n stands for "n" instances of S. In other words, S^n denotes the number "n". Any finite subset of **S** is clearly satisfied in $\mathbb{N}$. The Compactness theorem then says that there is a structure where all of the formulas in **S** are simultaneously true. This structure clearly cannot be $\mathbb{N}$ because then it would mean that there is an element in $\mathbb{N}$ that is larger than every natural number. This leads to the idea of structures for the axioms $N_1 - N_{10}$ with "infinite" elements.

The ideas that we have introduced here could be applied to other areas of mathematics, such as group theory, or set theory. You may find it curious that in order to talk about some of these results, we need an ambient set theory (to talk about the set $\mathbb{N}$, for example), and yet we can study set theory itself using the methods of mathematical logic. This is actually an important issue, the relation between logic and set theory. Set theory is itself a first order theory, where the variables range over sets. The non-logical axioms of set theory give us mechanisms to talk about subsets, power sets, infinite sets etc. And the resulting theory can be analyzed using logical methods and these logical methods rely on set theoretic intuitions. These issues are further discussed in [1]. We also refer to Kleene's book [2] for a treatment of some philosophical and technical issues that are not to be found in [1] and [3].

References

[1] W. Thomas H. D. Ebbinghaus, J. Flum, *Mathematical logic*, Springer Verlag, 1993, Though this is a lower-level text, this book has many advanced topics. I would recommend studying this book first if this article has interested you, and then looking at Shoenfield' book.

[2] Stephen C Kleene, *Introduction to metamathematics*, North Holland, 1967, This is another great classic in the subject; the book that Shoenfield says helped him write his classic. The style is old-fashioned, verbose, but the treatment is very precise, and it treats some topics such as intuitionism that the other books do not cover.

[3] Joseph R Shoenfield, *Mathematical logic*, reprinted by A.K. Peters, 2000, 1967, This book is a great classic in logic. It is very hard to read this book, but it has a wealth of information in it and many very challenging exercises. This book has played an important role in the education of many well-known logicians.

Ramanujan's Route to Roots of Roots

B. Sury

Stat-Math Unit, Indian Statistical Institute, 8th Mile Mysore Road, Bangalore 560 059, India

E-mail: sury@isibang.ac.in

Introduction

The fancy title[a] points out to a subject which originated with Ramanujan or, at least, one to which he gave crucial impetus to. We can't help but feel a sense of bewilderment on encountering formulae such as

$$\sqrt[3]{\sqrt[3]{2}-1} = \sqrt[3]{\frac{1}{9}} - \sqrt[3]{\frac{2}{9}} + \sqrt[3]{\frac{4}{9}};$$

$$\sqrt{\sqrt[3]{28}-\sqrt[3]{27}} = -\frac{1}{3}(-\sqrt[3]{98}+\sqrt[3]{28}+1);$$

$$\sqrt{\sqrt[3]{5}-\sqrt[3]{4}} = \frac{1}{3}(-\sqrt[3]{25}+\sqrt[3]{20}+\sqrt[3]{2});$$

$$\sqrt[3]{\cos\frac{2\pi}{7}} + \sqrt[3]{\cos\frac{4\pi}{7}} + \sqrt[3]{\cos\frac{8\pi}{7}} = \sqrt[3]{\frac{5-3\sqrt{7}}{2}};$$

$$\sqrt[6]{7\sqrt[3]{20}-19} = \sqrt[3]{\frac{5}{3}} - \sqrt[3]{\frac{2}{3}};$$

$$\sqrt{8-\sqrt{8+\sqrt{8-\cdots}}} = 1+2\sqrt{3}\sin 20^\circ;$$

$$\sqrt{23-2\sqrt{23+2\sqrt{23+2\sqrt{23-\cdots}}}} = 1+4\sqrt{3}\sin 20^\circ;$$

$$\frac{e^{-2\pi/5}}{1+}\,\frac{e^{-2\pi}}{1+}\,\frac{e^{-4\pi}}{1+}\,\frac{e^{-6\pi}}{1+}\cdots = \sqrt{\frac{5+\sqrt{5}}{2}} - \frac{\sqrt{5}+1}{2}.$$

The last expressions for the so-called Rogers–Ramanujan continued fraction appeared in Ramanujan's first letter to Hardy. These formulae are among some problems posed by

[a] Expanded version of a talk in IIT Madras on the occasion of the $\sqrt{1+119\sqrt{1+120\sqrt{1+121\sqrt{\cdots}}}}$-th birthday of someone who **radically** changed the mathematical landscape

Ramanujan in the Journal of the Indian Mathematical Society (see the article by Berndt, Choi & Kang in [R1]). As was usual with Ramanujan, he had a general formula hidden in the background and, singled out striking special cases in these problem sections. Usually, a study of his notebooks revealed what general formula he had in mind. The first thing to notice is that radicals are multi-valued and the meaning of expressions where radicals appear has to be made clear. Specially, where there is a 'nesting' of radicals, the level of complexity increases exponentially with each radical sign and it is computationally important to have equivalent expressions with the least number of radical signs. One more point to note is that often an equality, once written down, is almost trivial to verify simply by taking appropriate powers and simplifying. Thus, the intriguing question is to find out how such a formula was discovered and how to determine other such identities. The appropriate language to analyze this problem is the language of Galois theory.

1. Entry 4, page 108, 2nd notebook [R2]

Ramanujan posed the problem of finding the values of:

$$\sqrt{1+2\sqrt{1+3\sqrt{1+\cdots}}}$$

and

$$\sqrt{6+2\sqrt{7+3\sqrt{8+\cdots}}}.$$

These are special cases of Ramanujan's theorem appearing as Entry 4 on page 108, chapter 12 of his second notebook. Leaving aside the questions of convergence of these infinite radicals, the values can easily be discovered. Indeed, a beautiful, elementary discussion of the convergence and evaluation of these expressions appears in an article written by Shailesh Shirali ([S]). It is better to look at the more general form of the first expression

$$f(x)=\sqrt{1+x\sqrt{1+(x+1)\sqrt{1+(x+2)\sqrt{1+\cdots}}}}$$

for $x>0$. What is the meaning of this? One is looking to find functions $f:[0,\infty)\to[0,\infty)$ satisfying

$$f(x)=\sqrt{1+xf(x+1)}.$$

As $f(x)\geq 1$, we have

$$\begin{aligned} f(x) &\leq \sqrt{(1+x)f(x+1)} \\ &\leq \sqrt{1+x}\sqrt{\sqrt{(2+x)f(x+2)}}\leq\cdots \end{aligned}$$

Thus, $f(x)\leq\prod_{k=1}^{\infty}(k+x)^{\frac{1}{2^k}}$.

For $x\geq 1$, it is easy to bound the above infinite product above as

$$\prod_{k=1}^{\infty}(k+x)^{\frac{1}{2^k}}\leq\prod_{k=1}^{\infty}(2kx)^{\frac{1}{2^k}}=2x\prod_{k=1}^{\infty}k^{\frac{1}{2^k}}<2x\prod_{k=1}^{\infty}2^{\frac{k-1}{2^k}}\leq 4x.$$

In other words, for any $x\geq 0$, we have $f(x+1)<4(x+1)$ and, therefore, $f(x)=\sqrt{1+xf(x+1)}<\sqrt{1+4x(x+1)}=2x+1$. A fortiori, $f(x)<4x+1$ for any $x\geq 0$.

Playing the same game, if $f(x)\leq ax+1$ for some $a>0$ (and all $x\geq 0$), we get - on using $f(x+1)\leq a(x+1)+1$ - that

$$\begin{aligned} f(x) &\leq \sqrt{1+(a+1)x+ax^2} \\ &\leq \sqrt{1+(a+1)x+\left(\frac{a+1}{2}x\right)^2}\leq 1+\frac{a+1}{2}x. \end{aligned}$$

Hence, starting with $a=4$, we have the inequality $f(x)\leq ax+1$ recursively for $a=\frac{5}{2},\frac{7}{4},\frac{11}{8}$ etc. which is a sequence converging to 1. Thus, $f(x)\leq 1+x$ for all $x\geq 0$. Similarly, using the fact that $f(x+1)\geq f(x)$, one has $f(x)\geq\sqrt{1+xf(x)}$ which gives $f(x)\geq 1+\frac{x}{2}$ for $x\geq 0$. The earlier trick of iteration gives us that if $a>0$ satisfies $f(x)\geq 1+ax$ for all $x\geq 0$, then $f(x)\geq 1+\sqrt{a}x$. Thus, starting with $a=\frac{1}{2}$, we get $f(x)\geq 1+ax$ for $a=\frac{1}{2^{1/2^k}}$ for all $k\geq 1$. The latter sequence converges to 1 and we therefore have a perfect sandwich (a vegetarian version of which probably Ramanujan survived on, in England !) to get $f(x)=1+x$ for all $x\geq 0$. Therefore,

$$\sqrt{1+2\sqrt{1+3\sqrt{1+\cdots}}}=3(!)$$

We leave the reader to ponder about the value of

$$\sqrt{6+2\sqrt{7+3\sqrt{8+\cdots}}}.$$

2. A Theorem of Ramanujan

If m, n are arbitrary, then

$$\begin{aligned} \sqrt{m\sqrt[3]{4m-8n}+n\sqrt[3]{4m+n}} &= \pm\frac{1}{3}\big(\sqrt[3]{(4m+n)^2} \\ &\quad+\sqrt[3]{4(m-2n)(4m+n)}-\sqrt[3]{2(m-2n)^2}\big). \end{aligned}$$

As mentioned above, this is easy to verify simply by squaring both sides ! However, it is neither clear how this formula was arrived at nor how general it is. Are there more general formulae? In fact, it turns out that Ramanujan was absolutely on the dot here; the following result shows Ramanujan's result cannot be bettered:

Let $\alpha, \beta \in \mathbf{Q}^*$ such that α/β is not a perfect cube in Q. Then, $\sqrt{\sqrt[3]{\alpha}+\sqrt[3]{\beta}}$ can be denested if and only if there are integers m, n such that

$$\frac{\alpha}{\beta} = \frac{(4m-8n)m^3}{(4m+n)n^3}.$$

For instance, it follows by this theorem that $\sqrt{\sqrt[3]{3}+\sqrt[3]{2}}$ cannot be denested.

What is meant by denesting?

By the denesting of a nested radical one means rewriting it with fewer radical symbols. More formally, over any field K, nested radicals are defined as follows. Start with the elements in K – these are said to be nested radicals of depth zero. Use addition, subtraction, product, division and $\sqrt[n]{x}$ for $n \geq 2$ to form expressions (possibly not in K as we are taking n-th roots). With the expressions so formed, one can apply all the above procedures to form new expressions. Thus, a nested radical is an expression obtained from earlier-formed nested radicals by means of these procedures.

The usual convention used in fixing the values of radical expressions is as follows. An expression $\sqrt[3]{t}$ for a real number t will stand for the unique real cube root and, if s is a positive real number, $\sqrt{s}$ stands for the value which is the positive square root.

For example, the expression $\sqrt[3]{\sqrt{5}+2} - \sqrt[3]{\sqrt{5}-2}$ has value 1!

Indeed, if t is the value (according to the agreed-upon convention above), then t is seen to be a (real) root of the polynomial X^3+3X-4. As

$$X^3+3X-4 = (X-1)(X^2+X+4),$$

the only real root is 1.

3. Galois Theory for Denesting

To de-nest an expression such as $\sqrt{\sqrt[3]{\alpha}+\sqrt[3]{\beta}}$, one needs to locate a field where the de-nesting exists. Usually, one will need to go to a field where enough roots of unity exist. Let K be any field of characteristic zero and let $\bar{K}$ be an algebraic closure. The definition of depth over a field K implies that there are subfields $K^{(d)}$ of $\bar{K}$ defined by $\mathrm{depth}_K(x) = d$ for $x \in \bar{K}$ if and only if $x \in K^{(d)} \setminus K^{(d-1)}$. Here, $K^{(d)}$ is generated by radicals over $K^{(d-1)}$. In fact,

$$K^{(d)} := \{x \in \bar{K} : x^n \in K^{(d-1)} \text{for some } n\}.$$

For example,

$$\sqrt[6]{7\sqrt[3]{20}-19} = \sqrt[3]{\frac{5}{3}} - \sqrt[3]{\frac{2}{3}}$$

shows that the element on the left side which is in $\mathbf{Q}^{(2)}$ is actually contained in $\mathbf{Q}^{(1)}$ itself.

An element $x \in \bar{K}$ is a nested radical over K if and only if there exists a Galois extension L of K and a chain of intermediate fields

$$K \subset K_1 \subset \cdots \subset K_n = L$$

such that K_i is generated by radicals over K_{i-1} and $x \in L$.

Normally, if an element x is a nested radical over K, one obtains a chain as above successively generated by radicals such that $x \in L$ but L may not be automatically a Galois extension. For example, the left hand side above generates a non-Galois extension of $\mathbf{Q}$ and one needs to attach the 6-th roots of unity to get a Galois extension containing it.

So, why is it so important/useful to have a Galois extension?

The fact of the matter is that Galois's famous theorem tells us that $x \in \bar{K}$ is a nested radical if and only if the Galois closure of $K(x)$ over K has a solvable Galois group. Thus, the extensions $K^{(d)}$, if they are Galois extensions of $K^{(d-1)}$, have an abelian Galois group and this theory is well-studied under the title of 'Kummer theory'. Therefore, one may adjoin enough roots of unity at the first step of the chain to get a chain of Galois extensions and may apply Kummer theory. For instance,

$$\mathbf{Q} \subset \mathbf{Q}(\zeta_6) \subset \mathbf{Q}(\sqrt[3]{20}, \zeta_6) \subset \mathbf{Q}\left(\sqrt[6]{7\sqrt[3]{20}-19}, \zeta_6\right)$$

is a chain with each successive extension abelian.

Roughly speaking, Kummer theory (see [M]) can be summarized by:

Main Theorem of Kummer Theory. *If K contains the n-th roots of unity, then abelian extensions L of K whose Galois* *groups have exponent n correspond bijectively to subgroups* Ω *of* K^* *containing* $(K^*)^n$ *via* $L \mapsto K^* \cap (L^*)^n$ *and its inverse map* $\Omega \mapsto K(\Omega^{1/n})$.

Using Kummer theory, one may analyze all nested radicals x over $\mathbf{Q}$. In this talk, we do not go into the details of this general result proved by Mascha Honsbeek ([H]) in his doctoral thesis but focus mainly on the result relevant to Ramanujan's theorem. We will use the notation $\overline{\mathbf{Q}}$ for the field of all algebraic numbers. In this field, every nonconstant polynomial over it has all its roots in it. The following consequence of the above main theorem of Kummer theory and will be a key to denesting radicals.

Proposition 1. *Let* K *denote a field extension of* $\mathbf{Q}$ *containing the n-th roots of unity. Suppose* $a, b_1, b_2, \ldots, b_r \in \overline{\mathbf{Q}}^*$ *are so that* $a^n, b_1^n, \ldots, b_r^n \in K$. *Then,* $a \in K(b_1, \ldots, b_r)$ *if, and only if, there exist* $b \in K^*$ *and natural numbers* $m_1, m_2, \ldots, m_r$ *such that*

$$a = b\prod_{i=1}^{r} b_i^{m_i}.$$

Proof. The 'if' part is easily verified. Let us assume that $a \in L := K(b_1, \ldots, b_r)$. The subgroup Ω of L^* generated by the n-th powers of elements of (K^*) along with $b_1^n, \ldots, b_r^n$ satisfies $L = K(\Omega^{1/n})$ by Kummer theory. So, $a^n \in (L^*)^n \cap K^* = \Omega$. Thus, there exists $c \in K^*$ so that

$$a^n = c^n \prod_{i=1}^{r} b_i^{m_i n}.$$

Taking n-th roots on both sides and multiplying by a suitable n-th root of unity (remember they are in K), we get

$$a = b\prod_{i=1}^{r} b_i^{m_i}$$

for some $b \in K^*$. The proof is complete.

The following technical result from Galois theory which uses the above proposition is crucial in the denesting of $\sqrt{1 + \sqrt[3]{\beta/\alpha}}$ over $\mathbf{Q}$. Although this result is not difficult to prove, we shall not go into the details of the proof. We give a sketch for the cognoscenti; others may skip it.

Theorem 13. *Let c be a rational number which is not a perfect cube. Let* $\delta \in \mathbf{Q}(\sqrt[3]{c}) \setminus \mathbf{Q}$ *and let G denote the Galois group of the Galois-closure M of* $\mathbf{Q}(\sqrt{\delta})$ *over* $\mathbf{Q}$. *Then, the nested radical* $\sqrt{\delta}$ *can be denested over* $\mathbf{Q}$ *if, and only if, the second commutator group* G'' *of G is trivial. Further, these conditions are equivalent to the existence of* $f \in \mathbf{Q}^*$ *and some* $e \in \mathbf{Q}(\delta)$ *so that* $\delta = fe^2$.

Sketch of Proof. The essential part is to show that when G'' is trivial, then there are $f \in \mathbf{Q}^*$ and $e \in \mathbf{Q}(\delta)$ with $\delta = fe^2$.

We consider the field $K = \mathbf{Q}(\delta, \zeta_3)$, the smallest Galois extension of $\mathbf{Q}$ which contains δ. If δ_2, δ_3 are the other Galois-conjugates of δ in K, the **main claim** is that if $\delta_2\delta_3$ is not a square in K, then G'' is not trivial. To see this, suppose $\delta_2\delta_3$ (and hence its Galois-conjugates $\delta\delta_3, \delta\delta_2$) are non-squares as well. Then, $\sqrt{\delta\delta_2}$ cannot be contained in $K(\sqrt{\delta_2\delta_3})$ because of the above proposition. So, the extension $L = K(\sqrt{\delta\delta_2}, \sqrt{\delta_2\delta_3})$ has degree 4 over K and is contained in the Galois closure M of $\mathbf{Q}(\sqrt{\delta})$ over $\mathbf{Q}$. The Galois group Gal (L/K) is the abelian Klein 4-group V_4. Indeed, its nontrivial elements are $\rho_1, \rho_2, \rho_1\rho_2$ where ρ_1 fixes $\sqrt{\delta\delta_2}$ and sends $\sqrt{\delta_2\delta_3}$ and $\sqrt{\delta\delta_3}$ to their negatives; ρ_2 fixes $\sqrt{\delta_2\delta_3}$ and sends $\sqrt{\delta\delta_2}$ and $\sqrt{\delta\delta_3}$ to their negatives.

Also, Gal $(K/\mathbf{Q})$ is the full permutation group on $\delta, \delta_2, \delta_3$. We also put δ_1 instead of δ for convenience.

Suppose, if possible, $G'' = \{1\}$. Now, the second commutator subgroup of Gal $(L/\mathbf{Q})$ is trivial as it is a subgroup of G''. In other words, the commutator subgroup of Gal $(L/\mathbf{Q})$ is abelian.

Consider the action of Gal$(K/\mathbf{Q})$ on Gal (L/K) defined as:

$$(\sigma, \tau) \mapsto \sigma_L \tau \sigma_L^{-1}$$

where, for $\sigma \in$Gal $(K/\mathbf{Q})$, the element $\sigma_L \in$Gal $(L/\mathbf{Q})$ which restricts to K as σ.

The following computation shows that the commutator subgroup of Gal $(L/\mathbf{Q})$ cannot be abelian.

If π: Gal $(L/\mathbf{Q}) \rightarrow$ Gal $(K/\mathbf{Q})$ is the restriction map, look at any lifts a, b, c of (12), (13), (23) respectively. For any $d \in$ Gal (L/K), the commutator $ada^{-1}d^{-1}$ is defined independently of the choice of the lift a since Gal (L/K) is abelian. An easy computation gives:

$$a\rho_2 a^{-1}\rho_2^{-1} = \rho_1$$
$$b\rho_1 b^{-1}\rho_1^{-1} = \rho_1\rho_2$$
$$c(\rho_1\rho_2)c^{-1}(\rho_1\rho_2)^{-1} = \rho_2.$$

Therefore, the whole of Gal (L/K) is contained in the commutator subgroup of Gal $(L/\mathbf{Q})$. Now $(123) =$

(13)(23)(13)(23) implies that $d = bcb^{-1}c^{-1}$ which is in the commutator subgroup of Gal $(L/\mathbf{Q})$ is a lift of (123). Thus, $dgd^{-1}g^{-1} = Id$ for any $g \in$Gal (L/K) as Gal (L/K) is contained in the commutator subgroup of Gal $(L/\mathbf{Q})$ (an abelian group). But note that $d\rho_1 d^{-1}$ fixes $\sqrt{\delta_2\delta_3}$ and hence, cannot be equal to ρ_1. Thus, we have a contradiction to the assumption that $G'' = \{1\}$ while $\delta_2\delta_3$ is a non-square in K; the claim follows.

Now, assume that G'' is trivial. We would like to use the claim proved above to show that there are $f \in \mathbf{Q}^*$ and $e \in \mathbf{Q}(\delta)$ with $\delta = fe^2$.

Start with some $\eta \in K$ with $\delta_2\delta_3 = \eta^2$. We would like to show that $\eta \in \mathbf{Q}(\delta)$. This will prove our assertion, for then,

$$\delta = \frac{\delta_1\delta_2\delta_3}{\delta_2\delta_3} = \frac{\delta_1\delta_2\delta_3}{\eta^2} = fe^2$$

where $f = \delta_1\delta_2\delta_3 \in \mathbf{Q}$ and $e = \eta^{-1} \in \mathbf{Q}(\delta)$.

Suppose $\eta \notin \mathbf{Q}(\delta)$. Since the product $\delta_2\delta_3 \in \mathbf{Q}(\delta)$, on applying the above proposition to $K = \mathbf{Q}(\delta, \zeta_3) = \mathbf{Q}(\delta, \sqrt{-3})$, we get $\sqrt{\delta_2\delta_3} = \sqrt{-3}\theta$ for some $\theta \in \mathbf{Q}(\delta)$; that is,

$$\eta^2 = \delta_2\delta_3 = -3\theta^2.$$

Taking norms over $\mathbf{Q}$, we get $N(\eta)^2 = (-3)^3 N(\theta)^2$ which is a contradiction since $(-3)^3$ is not a square in $\mathbf{Q}$. Therefore, η indeed belongs to $\mathbf{Q}(\delta)$ and we are done.

4. Existence of De-nesting

In this section, we determine conditions under which elements e, f as in theorem 2 exist. For any non-zero α, β in $\mathbf{Q}$, the polynomial

$$F_{\beta/\alpha}(t) = t^4 + 4t^3 + 8\frac{\beta}{\alpha}t - 4\frac{\beta}{\alpha}$$

plays a role in determining the denestability of the nested radical $\sqrt{\sqrt[3]{\alpha} + \sqrt[3]{\beta}}$ over $\mathbf{Q}$.

Lemma 3. *Let $\alpha, \beta \in \mathbf{Q}^*$ such that α/β is not a perfect cube in $\mathbf{Q}$. Then, $\sqrt{\sqrt[3]{\alpha} + \sqrt[3]{\beta}}$ can be denested if and only if the polynomial $F_{\beta/\alpha}$ has a root in $\mathbf{Q}$.*

Proof. Now $\sqrt{\sqrt[3]{\alpha} + \sqrt[3]{\beta}}$ can be denested if and only if $\sqrt{1 + \sqrt[3]{\beta/\alpha}}$ can be denested. By the theorem, this happens if and only if there exists $f, x, y, z \in \mathbf{Q}$ with

$$1 + \sqrt[3]{\beta/\alpha} = f(x + y\sqrt[3]{\beta/\alpha} + z\sqrt[3]{\beta^2/\alpha^2})^2 \cdots\cdots \diamond$$

Assume that denesting can be done. The elements $1, \sqrt[3]{\beta/\alpha}, \sqrt[3]{\beta^2/\alpha^2}$ are linearly independent over $\mathbf{Q}$. Thus, we may compare like powers of $\sqrt[3]{\beta^2/\alpha^2}$ in $\diamond$ to get

$$1/f = x^2 + \frac{2yz\beta}{\alpha}$$
$$0 = y^2 + 2xz$$
$$1/f = \frac{\beta z^2}{\alpha} + 2xy$$

After a simple calculation, it is easy to see that $z \neq 0$ and that y/z is a root of $F_{\beta/\alpha}$.

Conversely, suppose $F_{\beta/\alpha}$ has a rational root s. Then, working backwards, a denesting is given as:

$$\sqrt{\sqrt[3]{\alpha} + \sqrt[3]{\beta}} = \pm\frac{1}{\sqrt{f}}(-\frac{s^2\sqrt[3]{\alpha^2}}{2} + s\sqrt[3]{\alpha\beta} + \sqrt[3]{\beta^2})$$

where $f = \beta - s^3\alpha$. The proof is complete.

Examples 4. *For $\alpha = 5, \beta = -4$ we get $s = -2$ to be the rational root of $F_{-4/5}(t) = t^4 + 4t^3 - \frac{32}{5}t + \frac{16}{5} = 0$. Thus, $f = -4 + 40 = 36$ and we have*

$$\sqrt{\sqrt[3]{5} - \sqrt[3]{4}} = \frac{1}{6}(-2\sqrt[3]{25} - 2\sqrt[3]{-20} + \sqrt[3]{16})$$
$$= \frac{1}{3}(-\sqrt[3]{25} + \sqrt[3]{20} + \sqrt[3]{2}).$$

Similarly, for $\alpha = 28, \beta = 27$, we have $s = -3$ and $f = 27^2$ and we get

$$\sqrt{\sqrt[3]{28} - \sqrt[3]{27}} = -\frac{1}{27}(-\frac{9}{2}\sqrt[3]{28^2} - 3\sqrt[3]{(-27)(28)} + \sqrt[3]{27^2})$$
$$= -\frac{1}{3}(-\sqrt[3]{98} + \sqrt[3]{28} + 1).$$

5. Connection with Ramanujan's De-nesting

We saw that denesting of $\sqrt{\sqrt[3]{\alpha} + \sqrt[3]{\beta}}$ involved the rational root of a certain related polynomial. The connection with Ramanujan's denesting comes while trying to characterize the α, β for which the polynomial $F_{\beta/\alpha}$ has a root in $\mathbf{Q}$. This is easy to see as follows:

Lemma 5. *Let $\alpha, \beta \in \mathbf{Q}^*$ where the ratio is not a cube. Then $\sqrt{\sqrt[3]{\alpha} + \sqrt[3]{\beta}}$ can be denested over $\mathbf{Q}$ if, and only if, $F_{\beta/\alpha}$ has a root s in $\mathbf{Q}$ which is if, and only if, there are integers m, n so that*

$$\frac{\alpha}{\beta} = \frac{(4m - 8n)m^3}{(4m + n)n^3}.$$

Proof. Of course, we need to prove only the second 'if and only if' and, even there, it suffices to prove the 'only if' part as the other implication is obvious. Now $s^4 + 4s^3 + 8s\beta/\alpha - 4\beta/\alpha = 0$ implies (on taking $s = n/m$) that

$$\frac{\beta}{\alpha} = \frac{s^3(s+4)}{4-8s} = \frac{(4m+n)n^3}{(4m-8n)m^3}.$$

Remarks. This is not quite the same as Ramanujan's denesting formula although the ratio β/α is the same. Ramanujan's theorem is:

$$\sqrt{m\sqrt[3]{4m-8n} + n\sqrt[3]{4m+n}} = \pm\frac{1}{3}\left(\sqrt[3]{(4m+n)^2} + \sqrt[3]{4(m-2n)(4m+n)} - \sqrt[3]{2(m-2n)^2}\right).$$

If we apply this to denest

$$\sqrt{\sqrt[3]{\alpha} + \sqrt[3]{\beta}}$$

where $\frac{\beta}{\alpha} = \frac{n^3(4m+n)}{m^3(4m-8n)}$, we would get

$$\sqrt{\sqrt[3]{\alpha} + \sqrt[3]{\beta}} = \frac{1}{\sqrt{m}}\sqrt[6]{\frac{\alpha}{4m-8n}}\sqrt{m\sqrt[3]{4m-8n} + n\sqrt[3]{4m+n}}.$$

This formula looks quite awkward and the natural question arises as to whether it is true that for integers α, β there exist integers m, n with $\alpha = m^3(4m - 8n)$ and $\beta = n^3(4m + n)$.

It turns out that this is not always the case. For example, if $\alpha = -4, \beta = 5$, the integers $m = n = 1$ work whereas for the other choice $\alpha = 5, \beta = -4$, there are no such integers (!) Thus, when we are denesting $\sqrt{\sqrt[3]{\alpha} + \sqrt[3]{\beta}}$, we are actually denesting $\sqrt{1 + \sqrt[3]{\beta/\alpha}}$ and it is better to use the method we discussed.

The asymmetry between m and n in Ramanujan's formula can be explained as follows. If one changes m to $m' = -\frac{n}{\sqrt{2}}$ and n to $n' = m\sqrt{2}$, it turns out that

$$4(m-2n)m^3 = (4m'+n')n'^3$$
$$(4m+n)n^3 = 4(m'-2n')m'^3.$$

Thus, we have the same denesting !

6. De-nesting Nested Square Roots

As mentioned in the beginning, using Galois theory, one may investigate the denesting of any nested radical in principle. However, it is not clear how to do in general. When the nested radical consists only of nested square roots, some nice results can be proved. We merely state two of them due to Borodin, Fagin, Hopcroft and Tompa ([BFHT]):

Theorem 6. *Let K have characteristic 0 and $a, b, r \in K$ with $\sqrt{r} \notin K$. Then,*

(a) $\sqrt{a + b\sqrt{r}} \in K(\sqrt{r}, \sqrt{a_1}, \ldots, \sqrt{a_n})$ *for some* $a_i \in K \Leftrightarrow \sqrt{a^2 - b^2 r} \in K$;

(b) $\sqrt{a + b\sqrt{r}} \in K(\sqrt[4]{r}, \sqrt{a_1}, \ldots, \sqrt{a_n})$ *for some* $a_i \in K \Leftrightarrow$ *either* $\sqrt{a^2 - b^2 r}$ *or* $\sqrt{r(-a^2 + b^2 r)}$ *is in* K.

Theorem 7. *Let $K \subset \mathbf{R}$ and $a, b, r \in K$ with $\sqrt{r} \notin K$. Let $a_1, \ldots, a_n \in K$ be positive and let $r_1, \ldots, r_n \geq 1$. If $\sqrt{a + b\sqrt{r}} \in K(\sqrt[r_1]{a_1}, \ldots, \sqrt[r_n]{a_n})$, then $\sqrt{a + b\sqrt{r}}$ is already in $K(\sqrt[4]{r}, \sqrt{b_1}, \ldots, \sqrt{b_n})$ for some $b_1, \ldots, b_n \in K$.*

These theorems also provide an algorithm for de-nesting. We do not go into it and proceed to the last section where we discuss certain values of the Rogers–Ramanujan continued fractions as nested radicals.

7. Rogers–Ramanujan Continued Fraction as Nested Radicals

The Rogers–Ramanujan continued fraction is the function

$$R(z) = \frac{q^{1/5}}{1+}\frac{q}{1+}\frac{q^2}{1+}\cdots$$

defined on the upper half plane (where $q = e^{2i\pi z}$). It is a holomorphic function and is also given by the product

$$R(z) = q^{1/5}\prod_{n=1}^{\infty}(1-q^n)^{\left(\frac{n}{5}\right)}$$

where $\left(\frac{n}{5}\right)$ here denotes the Legendre symbol. We had mentioned at the beginning Ramanujan's beautiful formula which he wrote in his first letter to Hardy:

$$\frac{e^{-2\pi/5}}{1+}\frac{e^{-2\pi}}{1+}\frac{e^{-4\pi}}{1+}\frac{e^{-6\pi}}{1+}\cdots = \sqrt{\frac{5+\sqrt{5}}{2}} - \frac{\sqrt{5}+1}{2}.$$

In other words, as $z = i$ gives $q = e^{-2\pi}$, we have

$$R(i) = \sqrt{\frac{5+\sqrt{5}}{2}} - \frac{\sqrt{5}+1}{2}.$$

Actually, this follows from a reciprocity theorem for $R(z)$ that Ramanujan wrote in his second letter to Hardy and which was proved by Watson. There are several other such identities written by Ramanujan like:

$$R(e^{i\sqrt{5}}) = \frac{\sqrt{5}}{1+\sqrt[5]{5^{3/4}\sqrt{(\frac{\sqrt{5}-1}{2})^5-1}}} - \frac{\sqrt{5}+1}{2}.$$

Each of these and many others can be proved using eta-function identities which were known to Ramanujan. Two such are:

$$\frac{1}{R(z)} - R(z) - 1 = \frac{\eta(z/5)}{\eta(5z)} \cdots\cdots\cdots (\heartsuit)$$

$$\frac{1}{R(z)^5} - R(z)^5 - 11 = \left(\frac{\eta(z)}{\eta(5z)}\right)^6 \cdots\cdots\cdots (\spadesuit)$$

Here, the Dedekind eta function $\eta(z) = q^{1/24}\prod_{n=1}^{\infty}(1-q^n)$ where $q = e^{2i\pi z}$ as before.

Note that $\eta(z)$ is essentially the generating function for the sequence $p(n)$ of partitions of a natural number n. Further, $\eta(z)^{24}$ is the discriminant function (sometimes called the Ramanujan cusp form)

$$\Delta(z) = q\prod_{n=1}^{\infty}(1-x^n)^{24}$$

which is the unique cusp form (up to scalars) of weight 12 and its Fourier expansion

$$\Delta(z) = (2\pi)^{12}\sum_{n=1}^{\infty}\tau(n)e^{2i\pi nz}$$

has Fourier coefficients $\tau(n)$, the Ramanujan tau function. Thus, this is familiar territory to Ramanujan. But, keeping in mind our focus on nested radicals, we shelve a discussion of proofs of these – there are proofs due to Berndt et al. in the spirit of Ramanujan and other ones by K. G. Ramanathan ([Ra]) using tools like Kronecker limit formula which were unfamiliar to Ramanujan. Readers interested in these may refer to [ABJL].

We go on to briefly discuss in the modern spirit how values of $R(z)$ at imaginary quadratic algebraic integers z on the upper half-plane can be expressed as nested radicals. This will involve class field theory, modular functions and the theory of complex multiplication. In order to be comprehensible to a sufficiently wide audience, we merely outline this important but rather technical topic.

What is the key point of this procedure?

The discussion below can be informally summed up as follows. The values of 'modular functions' (functions like $R(z)$) at an imaginary quadratic algebraic integer τ on the upper half-plane generate a Galois extension over $\mathbf{Q}(\tau)$ with the Galois group being abelian. This is really the key point. For, if we have any abelian extension L/K, one has a standard procedure whereby one could express any element of L as a nested radical over a field L' with degree $[L' : K] < [L : K]$. Recursively, this leads to a nested radical expression over K. Let us make this a bit more formal now.

Modular functions come in.

A meromorphic function on the extended complex upper half-plane $\mathcal{H} \cup \mathbf{P}^1(\mathbf{Q})$ which is invariant under the natural action of $\Gamma(N) := Ker(SL_2(\mathbf{Z}) \to SL_2(\mathbf{Z}/N\mathbf{Z}))$ is known as a *modular function of level N*. One has:

The Rogers-Ramanujan continued fraction $R(z)$ is modular, of level 5.

Now, a modular function of level N (since it is invariant under the transformation $z \mapsto z + N$) has a Fourier expansion in the variable $e^{2i\pi z/N} = q^{1/N}$. Those modular functions of level N whose Fourier coefficients lie in $\mathbf{Q}(\zeta_N)$ form a field F_N. In fact, in the language of algebraic geometry, F_N is the function field of a curve known as the modular curve $X(N)$ (essentially, the curve corresponding to the Riemann surface obtained by compactifying the quotient of the upper half-plane by the discrete subgroup $\Gamma(N)$). Moreover, F_N is a Galois extension of F_1 with Galois group $GL_2(\mathbf{Z}/N\mathbf{Z})/\{\pm I\}$. The classical modular j-function generates F_1 over $\mathbf{Q}$ and, hence, induces an isomorphism of $X(1)$ with $\mathbf{P}^1(\mathbf{Q})$. Let us briefly define it.

For τ on the upper half-plane, consider the lattice $\mathbf{Z} + \mathbf{Z}\tau$ and the functions

$$g_2(\tau) = 60\sum_{m,n}{}^{'}\frac{1}{(m+n\tau)^4}\left(= \frac{(2\pi)^4}{12}\left(1+\sum_{n=1}^{\infty}\sigma_3(n)e^{2\pi in\tau}\right)\right)$$

$$g_3(\tau) = 140\sum_{m,n}{}^{'}\frac{1}{(m+n\tau)^6}\left(= \frac{(2\pi)^6}{12}\left(1+\sum_{n=1}^{\infty}\sigma_5(n)e^{2\pi in\tau}\right)\right).$$

[Note that $p'(z)^2 = 4p(z)^3 - g_2(\tau)p(z) - g_3(\tau)$ where the Weierstrass p-function on $\mathbf{Z} + \mathbf{Z}\tau$ is given by $p(z) = \frac{1}{z^2} + \sum_{w}(\frac{1}{(z-w)^2} - \frac{1}{w^2})$.]

It can be shown that $\Delta(\tau) \stackrel{d}{=} g_2(\tau)^3 - 27g_3(\tau)^2 \neq 0$. The elliptic modular function $j : h \to \mathbf{C}$ is defined by

$$j(\tau) = 12^3 \cdot \frac{g_2(\tau)^3}{\Delta(\tau)}.$$

Similar to the generation of F_1 by the j-function is the fact that the field F_5 also happens to be generated over $\mathbf{Q}(\zeta_5)$ by a single function (in other words, $X(5)$ also has genus 0). An interesting classical fact is:

The function $R(z)$ generates F_5 over $\mathbf{Q}(\zeta_5)$.

The idea used in the proof of this is the following. The j-function can be expressed in terms of the eta function and on using the two identities ($\heartsuit$) and ($\spadesuit$), it follows that

$$j = \frac{(1 + 228R^5 + 494R^{10} - 228R^{15} + R^{20})^3}{(-R + 11R^6 + R^{11})}.$$

From this, one can derive also explicitly the minimal polynomial of R^5 over $\mathbf{Q}(j)$, a polynomial of degree 12. Thus, $\mathbf{Q}(R)$ has degree 60 over $\mathbf{Q}(j)$. In fact, this minimal polynomial is a polynomial in j with integer coefficients. In particular, if τ is an imaginary quadratic algebraic integer on the upper half-plane, $R(\tau)$ is an algebraic integer as this is true for j.

If τ is an imaginary quadratic algebraic integer on the upper half-plane, one can form the $\mathbf{Z}$-lattice $\mathcal{O}_\tau$ with basis $1, \tau$. This is an order (a subring containing a $\mathbf{Q}$-basis) in the imaginary quadratic field $\mathbf{Q}(\tau)$. When $\mathcal{O}_\tau$ is a maximal order, the 'class field' H of $\mathcal{O}_\tau$ is the 'Hilbert class field' of $\mathbf{Q}(\tau)$ - a finite extension of $\mathbf{Q}(\tau)$ in which each prime of $\mathbf{Q}(\tau)$ is unramified and principal. More generally, for any N, the field H_N generated by the values $f(\tau)$ of modular functions f of level N is nothing but the 'ray class field' of conductor N over $\mathbf{Q}(\tau)$ when $\mathcal{O}_\tau$ is a maximal order.

We have (see [L]):

The first main theorem of complex multiplication. *H_N is an abelian extension of* $\mathbf{Q}(\tau)$.

The Hilbert class field H is H_1 in this notation. That is, $H_1 = \mathbf{Q}(\tau)(j(\tau))$ is Galois over $\mathbf{Q}(\tau)$ with the Galois group isomorphic to the class group of $\mathcal{O}_\tau$. In general, for any N, one may explicitly write down the action of $\mathrm{Gal}(H_N/\mathbf{Q}(\tau))$ on $f(\tau)$ for any modular function f of level N. One may use that and the fact mentioned earlier that $R(\tau)$ is an algebraic integer to prove:

If τ is an imaginary quadratic algebraic integer on the upper half-plane, and $\mathcal{O}_\tau$ is the order in $\mathbf{Q}(\tau)$ with $\mathbf{Z}$-basis $1, \tau$, the ray class field H_5 of conductor 5 over $\mathbf{Q}(\tau)$ is generated by $R(\tau)$.

Let us begin with this data to describe an algorithm to express elements of H_N as nested radicals over $\mathbf{Q}(\tau)$.

Procedure for expressing as nested radicals

Let L/K be any abelian extension and let $w \in L$. Choose an element $\sigma \in Gal(L/K)$ of order $n > 1$. Take $L' = L^\sigma(\zeta_n)$ where L^σ denotes the fixed field under $< \sigma >$. Now, L' is an abelian extension of K and

$$\begin{aligned}[L' : K] &\leq \phi(n)[L^\sigma : K] < n[L^\sigma : K] \\ &= [L : L^\sigma][L^\sigma : K] = [L : K].\end{aligned}$$

Look at the Lagrange resolvents $h_i = \sum_{k=1}^n \zeta_n^{ik} w^{(\sigma^k)}$; $0 \leq i < n$ of w with respect to σ. Thus, we have

$$w = \frac{h_0 + h_1 + \cdots + h_{n-1}}{n}.$$

Since $h_0 = tr_{L/L^\sigma}(w) \in L'$, it is fixed by every element of $Gal(L(\zeta_n)/L')$. Furthermore, any element $\rho \in Gal(L(\zeta_n)/L')$ acts by some power of σ on L and as identity on ζ_n. Therefore, for $i = 1, 2, \ldots, n-1$;

$$h_i^\rho = \zeta_n^{-ia} h_i$$

where ρ is acting as σ^a on L.

Hence, we get $h_1^n h_2^n \cdots h_{n-1}^n \in L'$. Thinking of h_i as $\sqrt[m]{h_i^m}$, the expression

$$w = \frac{h_0 + h_1 + \cdots + h_{n-1}}{n}$$

is actually an expression as a nested radical over L'. Next, one can apply this procedure to $h_0, h_1^n, h_2^n, \ldots, h_{n-1}^n$ in L' etc.

Returning to our situation, when τ is an imaginary quadratic algebraic integer, we wish to look at the value $R(\tau)$. So, we will work with the ray class field H_5 over $\mathbf{Q}(\tau)$. Recall the identity

$$\frac{1}{R(z)} - R(z) - 1 = \frac{\eta(z/5)}{\eta(5z)} \cdots\cdots\cdots (\heartsuit)$$

It turns out that the elements $\frac{1}{R(\tau)}$ and $-R(\tau)$ of H_5 are Galois-conjugate over the field $\mathbf{Q}(\tau)(w(\tau))$ where we have written $w(\tau)$ for $\frac{\eta(\tau/5)}{\eta(5\tau)}$. Therefore, we have:

H_5 is generated over $\mathbf{Q}(\tau)$ by $w(\tau)$ and ζ_5.

Now $w(\tau)$ is an algebraic integer as both $\frac{1}{R(\tau)}$ and $-R(\tau)$ are so. Usually, one works with $w(\tau)$ instead of with $R(\tau)$ because the former has half the number of conjugates and also, it is defined in terms of the eta function which can be evaluated by software packages very accurately. The point is that the software packages exploit the fact that $SL_2(\mathbf{Z})$-transformations carry z to a new z with large imaginary part and so the Fourier expansion of $\eta(z)$ will converge quite rapidly. Therefore, to obtain expressions for $R(\tau)$, one works with $w(\tau)$ and uses ($\heartsuit$).

What τ to choose ?

Analyzing the product expression

$$R(z) = q^{1/5} \prod_{n=1}^{\infty} (1-q^n)^{\left(\frac{n}{5}\right)}$$

one observes that $R(z)$ is real when $Re(z) = 5k/2$ for some integer k. Thus, keeping in mind that we need algebraic integers τ with the corresponding order being maximal, we look at:

$$\tau_n = \sqrt{-n} \quad if \quad -n \not\equiv 1 \ mod \ 4;$$

$$\tau_n = \frac{5+\sqrt{-n}}{2} \quad if \quad -n \equiv 1 \ mod \ 4.$$

Observe that $\tau_1 = i$. In general, it can be shown that $\sqrt{5} \in \mathbf{Q}(\tau_n)$ and that $\frac{w(\tau_n)}{\sqrt{5}}$ is an algebraic integer. So, one can generate H_5 over $\mathbf{Q}(\tau_n)$ by $\tilde{w}(\tau_n) = \frac{w(\tau_n)}{\sqrt{5}}$ (instead of $w(\tau_n)$) along with ζ_5.

Applying the procedure of finding nested radicals in an abelian extension, one can show (using software packages):

$$\tilde{w}(\tau_1) = \tilde{w}(i) = 1.$$

Unwinding this for $w(i)$ and then for $R(i)$, one gets

$$R(i) = \sqrt{\frac{5+\sqrt{5}}{2}} - \frac{\sqrt{5}+1}{2} \ !$$

Similarly, one may obtain formulae like

$$R\left(\frac{5+i}{2}\right) = -\sqrt{\frac{5-\sqrt{5}}{2}} + \frac{\sqrt{5}-1}{2}$$

Last take on j-function

A romantic story goes that Ramanujan wrote somewhere that $e^{\pi\sqrt{163}}$ is 'almost' an integer ! Whether he actually did so or not, he could have known the following interesting fact.

Applying the first theorem of complex multiplication to $j(\tau)$ for τ imaginary quadratic, it follows that $j(\tau)$ is an algebraic integer of degree = class number of $\mathbf{Q}(\tau)$ i.e, $\exists$ integers $a_0, \ldots, a_{h-1}$ such that

$$j(\tau)^h + a_{h-1} j(\tau)^{h-1} + \cdots + a_0 = 0.$$

Now, there are only finitely many imaginary quadratic fields $\mathbf{Q}(\tau) = K$ which have class number 1. The largest D such that $\mathbf{Q}(\sqrt{-D})$ has class number 1 is 163. Since $163 \equiv 3(4)$, the ring of integers is $\mathbf{Z} + \mathbf{Z}(\frac{-1+i\sqrt{163}}{2})$. Thus $j(\frac{-1+i\sqrt{163}}{2}) \in \mathbf{Z}$.

The Fourier expansion of j has integer coefficients and looks like $j(\tau) = \frac{1}{q} + 744 + \sum_{n\geq 1} c_n q^n$ with $c_n \in \mathbf{Z}$ and

$$q = e^{2\pi i(\frac{\pm 1+i\sqrt{163}}{2})} = -e^{-\pi\sqrt{163}}.$$

Thus $-e^{\pi\sqrt{163}} + 744 - 196884\, e^{-\pi\sqrt{163}} + 21493760\, e^{-2\pi\sqrt{163}} + \cdots = j(\tau) \in \mathbf{Z}$. In other words,

$$e^{\pi\sqrt{163}} - integer = 196884\, e^{-\pi\sqrt{163}}$$
$$+\, 21493760\, e^{-2\pi\sqrt{163}} \ldots \approx 0!$$

To end, we should agree that Ramanujan **radically** changed the mathematical landscape!

References

[ABJL] G. E. Andrews, B. C. Berndt, L. Jacobsen & R. L. Lamphere, *The Continued Fractions found in the unorganized portions of Ramanujan's notebooks*, Memoir No. 477, Vol. 99 American Mathematical Society 1992.

[BFHT] A. Borodin, R. Fagin, J. E. Hopcroft & M. Tompa, *Decreasing the nesting depths of expressions involving square roots*, Journal of Symbolic computation, Vol. 1 (1985) pp.169–188.

[H] M. Honsbeek, Ph.D.Thesis, Universiteit Leiden 2005.

[L] S. Lang, *Elliptic functions*, GTM 112, Springer-Verlag 1987.

[M] P. Morandi, *Field and Galois theory*, GTM 167, Springer-Verlag 1996.

[R] K. G. Ramanathan, *Some applications of Kronecker's limit formula*, Journal of the Indian Mathematical Society, Vol. 52 (1987) pp. 71–89.

[R1] *Ramanujan: Essays and Surveys*, Edited by B. C. Berndt & R. A. Rankin, Culture and History of Mathematics, Vol. 1, 2001.

[R2] B. C. Berndt, *Ramanujan's Notebooks*, Springer-Verlag, New York 1994.

[S] S. Shirali, *An elementary problem that interested Ramanujan*, Resonance, Vol. 1, No. 12 (1996) pp. 60–62.