Theorem of Gelfond-Schneider on Transcendental Numbers

As an example of the application of complex analysis to number theory, we discuss the theorem of Gelfond and Schneider on transcendental numbers which was independently obtained in 1935 by Aleksandr Gelfond and Theodor Schneider. The theorem of Gelfond and Schneider solved the seventh problem of Hilbert which was presented to the 1900 International Congress of Mathematicians in Paris, which asks, "Is a^b transcendental, for algebraic $a \neq 0, 1$ and irrational algebraic b ?" The main tools are Jensen's formula and Siegel's lemma on solving for (algebraic) integer unknowns in homogeneous linear equations with (algebraic) integer coefficients by the use of the "box principle" (also known the "pigeon hole" principle).

Jensen's Formula. Let f be holomorphic on $\overline{\mathbb{D}_r}$ and $a_1, \cdots, a_n$ be the zeros of f in $\mathbb{D}$ (with multiplicities countered as repeated appearance) such that f is nonzero at 0. Then

$$\log|f(0)| = -\sum_{j=1}^{n} \log \frac{r}{|a_j|} + \frac{1}{2\pi} \int_{\theta=0}^{2\pi} \log|f(re^{i\theta})| d\theta.$$

It is a generalization of the mean value property of the harmonic function $\log|g|$ of a nowhere zero holomorphic function g. It is obtained by setting

$$g(z) = \frac{f(z)}{\prod_{k=1}^{n+p}(z - a_k)},$$

where $a_{n+1}, \cdots, a_{n+p}$ are the zeros of f on $\partial\mathbb{D}_r$ and applying the mean value property to the harmonic function $\log|g|$ on $\overline{\mathbb{D}_r}$, because

$$\int_{|z|=1} \log|z - a| \frac{dz}{z}$$

for $|a| = 1$.

When f vanishes to order s at 0, we have to replace $f(z)$ by $\frac{f(z)}{z^s}$ whose limit at $z = 0$ is $\frac{f^{(s)}(0)}{s!}$ and Jensen's formula becomes

$$\log\left|\frac{f^{(s)}(0)}{s!}\right| = -\sum_{j=1}^{n} \log \frac{r}{|a_j|} - s\log r + \frac{1}{2\pi} \int_{\theta=0}^{2\pi} \log|f(re^{i\theta})| d\theta,$$

because

$$\log\left|\frac{f(re^{i\theta})}{r^s e^{si\theta}}\right| = \log\left|f(re^{i\theta})\right| - s\log r.$$

Notations. For $X = (x_1, \cdots, x_n) \in K^n$, denote by $\|X\|$ the maximum of the absolute values of all conjugates of the coordinates x_i. We introduce the notion of *size*. For $B > 0$ an element α of K, the condition $\text{size}(\alpha) \leq B$ means that there exists a denominator d (which is a positive integer) for α such that $\log d \leq B$ and $\log \max_\sigma |\sigma\alpha| \leq B$. In other words,

$$\text{size}(\alpha) = \max(\log d, \log |\sigma\alpha|).$$

Siegel's Lemma. Given r homogeneous linear equations in n unknowns with integer coefficients of absolute value $\leq A$. Then there exists a nontrivial integer solution whose components have absolute value $\leq 2(2nA)^{\frac{r}{n-r}}$.

Proof. The coefficient matrix maps $\mathbb{Z}^n(B)$ into $\mathbb{Z}^r(nBA)$, where $\mathbb{Z}^n(B)$ is the subset defined by the absolute value of each coordinate no more than B. When $B^n > (nBA)^r$, two distinct vectors are mapped to the same. Their difference is a nontrivial solution. We can choose $B = (2nA)^{\frac{r}{n-r}}$. Then the absolute value of each component of the difference of two elements in $\mathbb{Z}^n(B)$ is no more than $\leq 2(2nA)^{\frac{r}{n-r}}$.

Coefficients of Equations in Number Field. Consider the case with coefficients in the ring I_K of algebraic integers in a number field K. An algebraic integer in K means an element of K which is the root of a *monic* polynomial with coefficients in $\mathbb{Z}$. Choose a $\mathbb{Z}$-basis $\omega_1, \cdots, \omega_M$ of I_K to transfer the r equations to rM equations with integer coefficients and nM unknowns. In the application of the box principle, the multiplication of the new $rM \times nM$ matrix from the old $r \times n$ matric by the unknown new nM-vector from the old n-vector involves the product $\omega_j\omega_k$. When the coefficients are elements α of K with a common denominator $d \in \mathbb{N}$ and with $\|\alpha\| \leq A$. There exists a nontrivial solution X in I_K such that $\|X\| \leq C_1(C_2ndA)^{\frac{r}{n-r}}$, where C_1, C_2 depend only on K.

Main Theorem. Let K be a number field (*i.e.,* K is an extension field of $\mathbb{Q}$ with $[K : \mathbb{Q}] < \infty$). Suppose $f(z), g(z)$ are entire functions of finite order $< \rho$ on $\mathbb{C}$ such that $f(z), g(z)$ are algebraically independent and satisfy the differential equation

$$\frac{df(z)}{dz} = P_1(f(z), g(z)) \quad \text{and} \quad \frac{dg(z)}{dz} = P_2(f(z), g(z))$$

for some polynomials $P_1(X, Y), P_2(X, Y) \in K[X, Y]$ with coefficients in K. The assumption that $f(z)$ is of finite order $< \rho$ means that $|f(z)| \leq C_\varepsilon e^{|z|^{\rho-\varepsilon}}$

for some sufficiently small $\varepsilon > 0$ and sufficiently large $C_\varphi > 0$. Let Z be the set of points in $\mathbb{C}$ such that both $f(z), g(z)$ are in K for $z \in Z$. Then the number of elements in the set Z is bounded by a constant $C_{\rho,K}$ which depends only on ρ and K.

Proof of Main Theorem. Let L be a natural number which will be allowed to go to infinity. Let J be the integral part of $\sqrt{L \log L}$. Choose m elements $z_0, z_1, \cdots, z_m$ of Z. Use Siegel's lemma to construct a polynomial $G(X,Y)$ with coefficients in K of degree $\leq J$ in each of the two variables X, Y such that the entire function $F(z) = G(f(z), g(z))$ vanishes at each of the points $z_0, z_1, \cdots, z_m$ to order $\geq L$. By relabelling $z_0, \cdots, z_m$, we assume that the vanishing order s of F at z_0 is the lowest so that $s \leq L$. To get a bound on m which depends only on ρ and K, we make a coordinate translation we assume without of generality that $z_0 = 0$.

We need to use Siegel's lemma to construct $G(X,Y)$ with appropriate bounds on its coefficients in K (which are the unknowns in the system of homogeneous linear equations) such that $G(f(z_j), g(z_j))$ vanishes to order $\geq L$ for $1 \leq j \leq m$. From the two differential equations we get

$$\|D^\lambda(f^j g^k)\| \leq (2J+L)^L C^{2J+L}$$

at points of Z for $\lambda \leq L$. To start out with, there is C^J from the absolute value of f, g at points of Z raised to power J. The other factors arise because of the use of the two differential equations. See the Appendix for the computational details. This means that in the application of Siegel's lemma we can choose

$$A = (2J+L)^L C^{2J+L}.$$

In the solution X from Siegel's lemma there is an exponent $\frac{r}{n-r}$ in the estimate of the solution

$$\|X\| \leq C_1 (C_2 n d A)^{\frac{r}{n-r}}.$$

For the time being, we use the notation r in the context of Siegel's lemma and not in the context of Jensen's formula. From $r \approx L$ and $n \approx J^2 \approx L \log L$, we have

$$\frac{r}{n-r} \approx \frac{1}{\log L}.$$

By "$\approx$" we mean the quotient of the two sides approaches 1 as $L \to \infty$ (or equivalently $J \to \infty$). From

$$\|X\| \leq C_1 (C_2 n d A)^{\frac{r}{n-r}} = C_1 \left(C_2 n d \left((2J+L)^L C^{2J+L} \right) \right)^{\frac{r}{n-r}}$$

it follows that

$$\log \|X\| \leq C_3 \frac{1}{\log L} (L \log L) \leq C_4 L.$$

All the constants C_j depend only on K and ρ.

Now we use the notation r in the context of Jensen's formula (and not in the context of Siegel's lemma). Choose $r \geq \max(|z_0|, |z_1|, \cdots, |z_m|)$. We now use Jensen's formula

$$(\ddagger) \qquad \begin{aligned} \sum_{j=1}^{m} \log \frac{r}{|z_j|} &\leq \sum_{j=1}^{m} \log \frac{r}{|z_j|} + s \log r \\ &\leq -\log \left| \frac{F^{(s)}(0)}{s!} \right| + \frac{1}{2\pi} \int_{\theta=0}^{2\pi} \log |F(re^{i\theta})| d\theta \\ &= \log(s!) - \log \left| F^{(s)}(0) \right| + \frac{1}{2\pi} \int_{\theta=0}^{2\pi} \log |F(re^{i\theta})| d\theta. \end{aligned}$$

Jensen's formula is replaced by an inequality, because $z_0, z_1, \cdots, z_m$ are not all the zeroes of $F(z)$ on $\{|z| < r\}$.

The size of the element of $-\log \left| F^{(s)}(0) \right|$ is no more than the sum of the size of the coefficients of $G(X, Y)$ and the size of the derivative $D^\lambda(f^j g^k)$ with $\lambda \leq J$. The size of the coefficients of $G(X, Y)$ is of the order $\leq L$ and the size of the *denominator* of the derivative $D^\lambda(f^j g^k)$ at $z_0 = 0$ with $\lambda \leq J$ is of order $\leq L$ as well. Here, when we say that an entity is of order $\leq L$ means that the entity is no more than a constant times L. From Stirling's formula

$$s! \sim \sqrt{2\pi} s^{s+\frac{1}{2}} e^{-s}$$

the order of growth of $\log(s!)$ is $s \log s \leq L \log L$.

We claim that in the inequality $(\ddagger)$ from the application of Jensen's formula, the expression

$$\log(s!) - \log \left| F^{(s)}(0) \right|$$

is no more than a universal constant times $[K : \mathbb{Q}] \, L \log L$. The factor $[K : \mathbb{Q}]$ enters from $-\log \left| F^{(s)}(0) \right|$ for the following reason. We have an estimate of the denominator of $F^{(s)}(0)$. The numerator of $F^{(s)}(0)$ is an algebraic integer but the absolute value of an algebraic integer may be < 1. The only thing we know is that its product with all its conjugates is a usual positive integer and therefore has absolute value ≥ 1 so that we have to add $([K:\mathbb{Q}]-1)$ times an upper bound for $\log\left|F^{(s)}(0)\right|$ in the estimate. This yields a term dominated by a constant times $[K:\mathbb{Q}]L\log L$.

Since the maximum of the absolute value of each of the points $z_0, z_1, \cdots, z_m$ is bounded independent of r, the contribution from the left-hand side of (‡) at least $mL\log r$ (because each vanishing order at $z_0, z_1, \cdots, z_m$ is at least L).

We now study the growth order of $\log\left|F(re^{i\theta})\right|$, to which both the coefficients a_{jk} of $P(X,Y)$ and $f^j g^k$ contribute. The size of a_{jk} is dominated by a constant times L. Since f and g are both of finite type $<\rho$, we know that for some sufficiently small $\varepsilon>0$,

$$\log\left|F(re^{i\theta})\right| \leq C_5(L+Jr^{\rho-\varepsilon}) \leq C_5(L+\sqrt{L\log L}\,r^{\rho-\varepsilon}).$$

In order to make the two terms on the right-hand side more or less of the same order, we choose to set $r=L^{\frac{1}{2\rho}}$ so that

$$\log\left|F(re^{i\theta})\right| \leq C_6\,L.$$

With our choices of the relative growth orders of J, L, and r, the final comparison of the growth order of the terms in (‡) is between

(i) the contribution $mL\log r=\frac{m}{2\rho}\,L\log L$ from $\sum_{j=1}^m \log\frac{r}{|z_j|}$ on the left-hand side and

(ii) the contribution of the order $[K:\mathbb{Q}]\,L\log L$ from $\log(s!)-\log\left|F^{(s)}(0)\right|$, plus the contribution of order L from

$$\frac{1}{2\pi}\int_{\theta=0}^{2\pi}\log|F(re^{i\theta})|d\theta$$

on the right-hand side.

So the final answer is

$$m\leq C_7\rho\,[K:\mathbb{Q}].$$

Q.E.D.

Application to Transcendence of e and π and the Seventh Problem of Hilbert. The above argument can be applied to more than two meromorphic functions to give the following theorem.

Theorem (in More General Setting). Let K be a finite extension of the rational numbers. Let $f_1, \cdots, f_N$ be meromorphic functions of order $\leq \rho$. Assume that the field $K(f_1, \cdots, f_N)$ has transcendence degree ≥ 2 over K and that the derivative $D = d/dz$ maps the ring $K[f_1, \cdots, , f_N]$ into itself. Let $w_1, \cdots, , w_m$ be distinct complex numbers not lying among the poles of the f_i, such that

$$f_i(w_v) \in K, \text{ for all } i = 1, \cdots, N \text{ and } v = 1, \cdots, m.$$

Then $m \leq 20\rho\,[K : Q]$.

(See Theorem 1 on p.21 of Lang's book of 1966 entitled "Introduction to Transcendental Numbers".)

Here are some applications of our Main Theorem, which yield the transcendence of e and π and the solution of the Seventh Problem of Hilbert by Gelfond and Schneider in 1935.

Corollary 1 (Hermite-Lindemann). If α is algebraic (over $\mathbb{Q}$) and $\alpha \neq 0$, then e^{α} is transcendental. Hence e and π are transcendental.

Proof. Suppose α and e are algebraic. Let $K = \mathbb{Q}(\alpha, e^{\alpha})$. The functions z and e^z are algebraically independent over K since if e^z is the root of some polynomial with coefficients in $K(z)$

$$(e^z)^n + a_{n-1}(z)(e^z)^{n-1} + \cdots + a_1(z)e^z + a_0(z) = 0,$$

then the term $(e^z)^n = e^{nz}$ on the left dominates all the other terms for large z, which contradicts the above equality. The ring $K[z, e^z]$ is obviously mapped into itself by the derivative. Now for any $m \geq 1$ we can set

$$w_1 := \alpha, w_2 := 2\alpha, \cdots, w_m := m\alpha,$$

and our functions z and e^z take on algebraic values at all w_i. This leads to the contraction that m is bounded by a constant which depends only on the order $\rho = 1$ of e^z and K.

Since $e^1 = e$ and $e^{2\pi i} = 1$, it follows that e and π are transcendental. Q.E.D.

Corollary 2 (Solution of the Seventh Problem of Hilbert by Gelfond-Schneider in 1935). If α is algebraic, $\alpha \neq 0, 1$ and if β is algebraic and irrational, then $\alpha^\beta = e^{\beta \log \alpha}$ is transcendental.

Proof. We proceed as in proving Corollary 1, but now we consider the functions $e^{\beta z}$ and e^z. If they are algebraically dependent then $e^{\beta z}$ and e^z would be the roots of a polynomial $q(T_1, T_2)$, so

$$0 = q\left(e^{\beta z}\right) = \sum_{i,j=0}^{N} b_{ij} \left(e^{\beta z}\right)^i (e^z)^j = \sum_{i,j=0}^{N} b_{ij} e^{(i\beta + j)z}.$$

For this equation to hold we must have cancellations of two or more terms, *i.e.,* for some i_1, i_2 and j_1, j_2 we have $i_1\beta + j_1 = i_2\beta + j_2$, or

$$(i_1 - i_2)\beta = j_2 - j_1.$$

This implies that either $i_1 = i_2$ and $j_1 = j_2$ or β is rational. Now let

$$w_1 := \log \alpha, w_2 := 2 \log \alpha, , \cdots , w_m := m \log \alpha,$$

so our functions e^z and $e^{\beta z}$ take on algebraic values at the w_i. This yields the same kind of contradiction.

APPENDIX: Estimation of Derivatives

The purpose of this Appendix is to introduce a system of notations so that we can more effortlessly drive

$$\|D^\lambda(f^j g^k)\| \leq (2J+L)^L C^{2J+L} \quad \text{at points of } Z \ \text{ for } \lambda \leq L$$

from

$$\frac{df(z)}{dz} = P_1(f(z), g(z)) \quad \text{and} \quad \frac{dg(z)}{dz} = P_2(f(z), g(z)).$$

System of Notations. Let $T_1, \cdots, T_N$ be variables and

$$P(T_1, \cdots, T_N) = \sum \alpha_\nu T_1^{\nu_1} \cdots T_N^{\nu_N}$$

with $\alpha_\nu \in \mathbb{C}$. We define $|P| = \max |\alpha_\nu|$. For

$$Q(T_1, \cdots, T_N) = \sum \beta_\nu T_1^{\nu_1} \cdots T_N^{\nu_N}$$

with $\beta_\nu \geq 0$ we say that $P \prec Q$ if $|a_\nu| < \beta_\nu$. clearly $P_i \prec Q_i$ implies $P_1 + P_2 \prec Q_1 + Q_2$ and $P_1 P_2 \prec Q_1 Q_2$ and $D_i P \prec D_i Q$, where D_i means differentiation with respect to T_i. If the total degree of P is $\leq r$, then

$$P \prec |P|(1 + T_1 + \cdots + T_N)^r.$$

Suppose K is a number field. We denote by $\|\cdot\|$ the maximum of all absolute values. When the coefficients of P belong to K we can define $\|P\|$ in the same way that $|P|$ is defined. The denominator for P means the common denominator for the coefficients of P.

Proposition. Suppose $f_1, \cdots, f_N$ are holomorphic functions in a neighborhood of w in $\mathbb{C}$. Assume that $D = \frac{d}{dw}$ maps $K[f_1, \cdots, f_N]$ to itself. Assume that $f_i(w) \in K$ for $1 \leq i \leq N$. Assume that there exists $P_i(T_1, \cdots, T_N)$ such that the coefficients of P_i are in K and

$$Df_i = P_i(f_1, \cdots, f_N).$$

Let δ be the maximum degree of $P_1, \cdots, P_N$. Then

$$\|D^k f(w)\| \leq \|P\| r^k k! C_1^{k+r},$$

where δ depends on δ. Furthermore, there is a denominator for $D^k f(w)$ bounded by $\mathrm{den}(P) C_1^{k+r}$.

Proof. We assume that

$$DT_i = P_i(T_1, \cdots, T_N).$$

By the chain rule,

$$DP(T_1, \cdots, T_N) = \sum_{i=1}^{N} (D_i P)(T_1, \cdots, T_N) P_i(T_1, \cdots, T_N)$$

and

$$(D_i P)(T_1, \cdots, T_N) \prec D_i \left(|P|(1 + T_1 + \cdots + T_N)^r \right) = r|P|(1+T_1+\cdots+T_N)^{r-1}.$$

This means

$$\begin{aligned} DP(T_1, \cdots, T_N) &\prec \sum_{i=1}^{N} r|P|(1 + T_1 + \cdots + T_N)^{r-1} |P_i|(1 + T_1 + \cdots + T_N)^{\delta} \\ &\prec \left(\sum_{i=1}^{N} |P_i| \right) |P| r (1 + T_1 + \cdots + T_N)^{r+\delta-1} \end{aligned}$$

and

$$D^k P \prec \left(\sum_{i=1}^{N} |P_i| \right)^k |P| r (r+(\delta-1)) \cdot (r+2(\delta-1)) \cdot \cdots \cdots (r+k(\delta-1))(1+T_1+\cdots+T_N)^{r+k(\delta-1)}.$$

Now plug in the value $T_\nu = f_\nu(w)$. We get a constant C_1 such that

$$\|D^k f(w)\| \leq \|P\| r^k k! C_1^{k+r}.$$

Furthermore, there is a denominator for $D^k f(w)$ bounded by $\mathrm{den}(P) C_1^{k+r}$. Q.E.D.