Chapter Five
Schneider's Solution to Hilbert's Seventh Problem (and Beyond)

(Revised January 2, 2011)

In this lecture we will briefly examine Schneider's solution to Hilbert's Seventh Problem and then look at two alternate, more modern, approaches to the last part of this proof (wherein we obtain the nonzero algebraic number that leads to the ultimate contradiction). We begin with Schneider's construction of the auxiliary function which requires an application of Siegel's Lemma from the last chapter. For clarity we restate the proposition implicit in Schneider's solution.

Proposition. *Suppose α and β are algebraic numbers with $\alpha \neq 0, 1$ and β irrational. Further assume that α^β is algebraic and let $d = [\mathbf{Q}(\alpha, \beta, \alpha^\beta) : \mathbf{Q}]$. Let m be a positive integer and put $D_1 = [\sqrt{2d}m^{3/2}]$ and $D_2 = [\sqrt{2d}m^{1/2}]$. Then if m is sufficiently large there exist rational integers $c_{k\ell}, 0 \leq k \leq D_1 - 1,\ 0 \leq \ell \leq D_2 - 1$, not all zero, such that the function*

$$F(z) = \sum_{k=0}^{D_1-1} \sum_{\ell=0}^{D_2-1} c_{k\ell} z^k \alpha^{\ell z} \tag{1}$$

satisfies

$$F(a + b\beta) = 0 \text{ for } 1 \leq a, b \leq m. \tag{2}$$

Moreover, there exists a constant $c_0 = c_0(\alpha, \beta)$ so that the integers $|c_{k\ell}|$ satisfy

$$0 < \max |c_{k\ell}| \leq c_0^{m^{2/3} \log m}. \tag{3}$$

Note: Siegel's Lemma enabled Schneider to describe the function he desired for his proof. That lemma does not explicitly yield the integral solutions $X_1, \ldots, X_N$, it only establishes that they exist and that they satisfy (3). However, Schneider would know that for any pairs of integers a and b

$$\begin{aligned} F(a + b\beta) &= \sum_{k=0}^{D_1-1} \sum_{\ell=0}^{D_2-1} c_{k\ell}(a + b\beta)^k e^{\ell \log \alpha (a+b\beta)} \\ &= \sum_{k=0}^{D_1-1} \sum_{\ell=0}^{D_2-1} c_{k\ell}(a + b\beta)^k \alpha^{a\ell} \alpha^{\beta b\ell}, \end{aligned}$$

is an integral polynomial expression involving $\alpha, \beta,$ and α^β and so is an algebraic number.

There is one small twist to applying Siegel's Lemma to obtain the appropriate function–the coefficients in Schneider's system of equations are not rational integers, as are required by the lemma in order to obtain integral solutions, but algebraic numbers. This does not present too great of an obstacle; we simply represent each algebraic number in terms of a primitive element of a field containing all of the algebraic numbers under consideration, and then set each coefficient equal to zero.

Outline of the proof of this Proposition

Step 1. Translate the condition that $F(a+b\beta)=0$ for $1 \le a, b \le m$ into a system of m^2 linear equations with *algebraic coefficients*.

Step 2. Using a primitive element for the number field $K = \mathbf{Q}[\alpha, \beta, \alpha^\beta]$, translate the condition $F(a+b\beta)=0$ for $1 \le a, b \le m$ into a larger system of equations with rational integral coefficients.

Step 3. Apply Siegel's Lemma to obtain the appropriate function $F(z)$.

Details of the proof.

Step 1. We begin by representing our desired function $F(z)$ with indeterminant coefficients $c_{k\ell}$ and unspecified degrees D_1 and D_2 :

$$F(z) = \sum_{k=0}^{D_1-1} \sum_{\ell=0}^{D_2-1} c_{k\ell} z^k e^{(\log\alpha)\ell z}.$$

In order to translate the vanishing of $F(z)$ at all of the desired points $a+b\beta$ into a rather explicit homogeneous system of linear equations with integer coefficients it helps to introduce informative notation for the coefficients of the system of linear equations corresponding to the conditions

$$F(a+b\beta)=0 \quad \text{for} \quad 1 \le a \le m, 1 \le b \le m \ .$$

Since

$$\begin{aligned} F(a+b\beta) &= \sum_{k=0}^{D_1-1} \sum_{\ell=0}^{D_2-1} c_{k\ell}(a+b\beta)^k e^{(\ell\log\alpha)(a+b\beta)} \\ &= \sum_{k=0}^{D_1-1} \sum_{\ell=0}^{D_2-1} \underbrace{(a+b\beta)^k(\alpha)^{\ell a}(\alpha^\beta)^{\ell b}}_{\text{coefficients}} \ \underbrace{c_{k\ell}}_{\text{unknowns}} = 0 \ , \end{aligned}$$

we see that for each choice of integers k, ℓ, a, b we need to understand the algebraic number

$$(a+b\beta)^k e^{(\log\alpha)\ell a} e^{(\beta\log\alpha)\ell b} = (a+b\alpha)^k(\alpha)^{\ell a}(\alpha^\beta)^{\ell b}. \tag{4}$$

Step 2. Let θ be a primitive element for the field $K = \mathbf{Q}(\alpha, \beta, \alpha^\beta)$, where $d = [K : \mathbf{Q}]$, which is also an algebraic integer. Then there are rational polynomials p_α, p_β and p_{α^β} so that

$$\alpha = p_\alpha(\theta), \beta = p_\beta(\theta) \text{ and } \alpha^\beta = p_{\alpha^\beta}(\theta).$$

Then a typical term in the summand representing $F(a+b\beta)$ may be rewritten as:

$$(a+b\beta)^k e^{(\log\alpha)\ell a} e^{(\beta\log\alpha)\ell b} = (a+bp_\beta(\theta))^k (p_\alpha(\theta))^{\ell a} (p_{\alpha^\beta}(\theta))^{\ell b},$$

We let $\delta = den(p_\alpha(\theta))den(p_\beta(\theta))den(p_{\alpha^\beta}(\theta))$. Then

$$\delta^{D_1+2D_2m}F(a+b\beta)$$

may be rewritten to only involve algebraic integers. And although the above expression will involve powers of θ greater than $d-1$, each of these may be rewritten as a linear combination of $1,\theta,\dots,\theta^{d-1}$ with coefficients of predictable absolute values. (See for example the exercises at the end of this chapter). In particular a typical summand in $\delta^{D_1+2D_2m}F(a+b\beta)$ may be rewritten as,

$$\begin{aligned}\delta^{(D_1-k)+(D_2m-\ell a)+(D_2m-\ell b)}&(\delta(a+bp_\beta(\theta)))^k(\delta p_\alpha(\theta))^{\ell a}(\delta p_{\alpha^\beta}(\theta))^{\ell b}\\ &= a_1(k,\ell,a,b)+a_2(k,\ell,a,b)\theta+\ldots+a_d(k,\ell,a,b)\theta^{d-1} \quad (5)\end{aligned}$$

where the integers $a_1,a_2,\dots,a_d$ satisfy

$$\max_{1\le r\le d}|a_r(k,\ell,a,b)| \le c_1^{D_1\log m+D_2m}, \quad (6)$$

c_1, and the other constants $c_2,\dots$ below, depend only on α,β, and our choice of θ, but not any of the parameters.

Thus pulling all our observations together, we see that for each pair of integers (a,b), we have

$$\delta^{D_1+2D_2m}F(a+b\beta) = A_1+A_2\theta+\cdots+A_d\theta^{d-1} \ ,$$

where each integer $A_j = A_j(a,b)$ can be expressed as

$$A_j = \sum_{k=0}^{D_1-1}\sum_{\ell=0}^{D_2-1} t_j(k,\ell,a,b)c_{k\ell} \ ,$$

and each coefficient $t_j(k,\ell,a,b)$ is an *integer* arising from a sum of D_1D_2 integers $a_j(k,\ell,a,b)$. Thus given our previous upper bounds, it follows that each coefficient satisfies

$$|t_j(k,\ell,a,b)| \le D_1D_2c_2^{D_1\log m+D_2m} \ .$$

(We note for later use that in obtaining the above estimate we used the bounds $1\le a,b\le m$. We will later need to appeal to the more explicit upper bound:

$$|t_j(k,\ell,a,b)| \le D_1D_2c_2^{D_1\log\max\{|a|,|b|\}+D_2\max\{|a|,|b|\}}) \ .$$

Since the numbers $1,\theta,\theta^2,\dots,\theta^{d-1}$ are $\mathbf{Q}$-linearly independent, it follows that $\delta^{D_1+2D_2m}F(a+b\beta)=0$, so $F(a+b\beta)$ equals 0, if and only if each of the associated quantities $A_1,A_2,\dots,A_d$ equals 0. Therefore we can replace each single linear equation $F(a+b\beta)=0$ involving algebraic coefficients with d linear equations involving only integer coefficient. Namely,

$$A_1(a,b)=0\ ,\quad A_2(a,b)=0\ ,\ldots,\ A_d(a,b)=0\ .$$

Step 3. For each pair (a,b), if we set each of the associated linear forms $A_1, A_2, \ldots, A_d$ equal to 0, then we obtain a homogeneous system of dm^2 linear equations in D_1D_2 unknowns. By Siegel's Lemma, if

$$D_1D_2 > dm^2\ ,$$

then there exist integers $c_{k\ell}$, not all zero, that form a solution to the linear system

$$A_j(a,b)=\sum_{k=0}^{D_1-1}\sum_{\ell=0}^{D_2-1} t_j(k,\ell,a,b)c_{k\ell}=0\ ,$$

for $j=1,2,\ldots,d$, $a=1,\ldots,m$, and $b=1,\ldots,m$, such that for each m and n,

$$|c_{k\ell}| < \left((D_1D_2)^2 c_2^{D_1\log m + D_2 m}\right)^{\frac{dm^2}{D_1D_2-dm^2}}\ .$$

In order to simplify this upper bound we can now fix a relationship between the parameters D_1, D_2, and m. The natural thing to try is to balance the two exponents in (6), i.e., take $D_1\log m$ equal to D_2m. The inclusion of a logarithmic term is necessary in many transcendence proofs, for example in Gelfond's solution to Hilbert's seventh problem, but in Schneider's somewhat less delicate proof we can ignore the relatively slow-growing $\log m$ factor. We choose D_1 and D_2 such that

$$D_1D_2 = 2dm^2 \ \text{ and } \ D_1 = D_2m\ .$$

If we imagine that m is our free parameter, we solve for D_1 and D_2 and obtain:

$$D_1=\sqrt{2d}\,m^{3/2} \ \text{ and } \ D_2=\sqrt{2d}\,m^{1/2}\ ,$$

with the additional understanding that we will henceforth take m such that these quantities are integers (i.e., take m always of the form $m=2dn^2$ where n is an integer).

We note that indeed $D_1D_2=2dm^2>dm^2$ as required in Siegel's Lemma. So applying Siegel's Lemma we see that for m sufficiently large there exist integers $c_{k\ell}$, not all zero, satisfying

$$|c_{k\ell}| < c_0^{m^{3/2}\log m}\ , \tag{7}$$

so that if $P(x,y)=\sum_{k=0}^{D_1-1}\sum_{\ell=0}^{D_2-1} c_{k\ell}x^ky^\ell$, and $F(z)=P(z,e^{\log\alpha z})$, then $F(z)$ is a nonzero function with the property that for each $a=1,\ldots,m$ and $b=1,\ldots,m$,

$$F(a+b\beta)=0\ .$$

This completes the proof of the proposition.

Once we have a function with prescribed zeros we need a nonzero value of the function that leads to a nonzero algebraic integer whose norm is less than 1. This requires three things: a nonzero value of the function, an upper bound for the absolute value of this nonzero value, and information about the nonzero value's conjugates. An important observation that will assist us in meeting all of these requirements is that since β is irrational, we see that $a + b\beta = a' + b'\beta$ if and only if $a = a'$ and $b = b'$. Therefore, by our construction, $F(z)$ has at least m^2 *distinct* zeros, namely, at $z = a + b\beta$, for $1 \le a \le m$ and $1 \le b \le m$.

The conclusion of the proof.

Before we discuss *how* to find a nonzero value for the function $F(z)$ let's examine how such a nonzero value leads to a completion of the proof. We will return to guarantee the existence of an appropriate nonzero value in the next section.

Using algebraic conjugates. Using a fairly complicated determinant argument, see below, Schneider proved that there exists a pair of integers a^*, b^*, with $1 \le a^*, b^* < 4m$ so that $F(a^* + b^*\beta) \ne 0$. In this section we use the algebraic norm to obtain a nonzero integer from the nonzero algebraic number $F(a^*+b^*\beta)$ whose absolute value is less than 1. Although we have already seen such an argument in some detail we will give fairly complete details here. However we will not explicitly display the dependence on $\alpha, \beta,$ or α^β, and consequently on θ and d, in our estimates. Rather we continue to absorb these explicit dependencies into consecutively numbered constants $c_1, c_2, \ldots$.

We begin by letting $m^* = \min\{a^*, b^*\}$, with the property that

$$F(a + b\beta) = 0 \text{ for } 1 \le a, b < m^* \quad \text{and } F(a^* + b^*\beta) \ne 0.$$

As we will see in the next section Schneider showed that $m \le m^* < 4m$.

Recalling our earlier notation if we recompute the estimates required to apply Siegel's Lemma using the specific numbers a^* and b^*, which give rise to a nonzero value for the function $F(z)$, instead of with general a and b satisfying $1 \le a, b \le m$ we have

$$\delta^{D_1+2D_2m^*} F(a^* + b^*\beta) = A_1^* + A_2^*\theta + \cdots + A_d^*\theta^{d-1},$$

where the integers $A_j^* = A_j^*(a^*, b^*)$ satisfy

$$\begin{aligned}\max_{1\le j\le d}\{|A_j^*|\} &\le D_1D_2\delta^{D_1+2D_2m^*} \max\{|c_{k\ell}|\} \max\{|t_j(k,\ell,a^*,b^*)|\} \\ &\le D_1D_2c_3^{D_1+D_2m^*} c_4^{m^{3/2}\log m} c_5^{D_1\log m^*+D_2m^*} \\ &\le c_6^{m^{3/2}\log m}, \text{ since } m^* \le 4m.\end{aligned}$$

We let $\theta_1 = \theta, \theta_2, \ldots, \theta_d$ denote the conjugates of θ and, simplifying notation a bit, consider the product

$$N = \underbrace{\left(A_1^* + A_2^*\theta + \cdots + A_d^*\theta^{d-1}\right)}_{\text{primary factor}} \underbrace{\prod_{i=2}^{d} \left(A_1^* + A_2^*\theta_i + A_3^*\theta_i^2 + \cdots + A_d^*\theta_i^{d-1}\right)}_{\text{secondary factors}} .$$

This product is a nonzero rational integer since $\delta^{D_1+2D_2m}F(a^* + b^*\beta)$ is a nonzero algebraic integer, i.e., $N \neq 0$.

Since the argument leading to an upper bound for $|N|$ is so similar to the argument we used to conclude Gelfond's proof we will be brief. We estimate the absolute value of the *primary factor* through an application of the Maximum Modulus Principle; this estimate depends in a crucial way on the number of zeros of the function $F(z)$. The absolute value of each of the *secondary factors* is estimated through a simple application of the triangle inequality (given the above estimate for $\max\limits_{1\le j\le d}\{|A_j^*|\}$, above).

Estimating the primary factor. To estimate the primary factor $A_1^* + A_2^*\theta + A_3^*\theta^2 + \cdots + A_d^*\theta^{d-1}$ we apply the Maximum Modulus Principle to the function

$$G(z) = \frac{\delta^{D_1+2D_2m^*}F(z)}{\prod_{a=1}^{m-1}\prod_{b=1}^{m-1}(z-(a+b\beta))} .$$

$G(z)$ is an entire function and

$$|\delta^{D_1+2D_2m^*}F(a^* + b^*\beta)| = |G(a^* + b^*\beta)| \prod_{a=1}^{m-1}\prod_{b=1}^{m-1} |(a^* - a) + (b^* - b)\beta| .$$

Given that $a \le m$ and $b \le m$, we have that for any $R > m^*(1+|\beta|)$, $|a^*+b^*\beta| < R$, so the Maximum Modulus Principle implies that

$$\begin{aligned} |\delta^* F(a^* + b^*\beta)| &\le |G|_R \prod_{a=1}^{m-1}\prod_{b=1}^{m-1} |(a^* - a) + (b^* - b)\beta| \\ &\le \frac{|\delta^{D_1+2D_2m^*}F|_R}{\left|\prod_{a=1}^{m-1}\prod_{b=1}^{m-1}(z-(a+b\beta))\right|_R} \prod_{a=1}^{m-1}\prod_{b=1}^{m-1} |(a^* - a) + (b^* - b)\beta| . \end{aligned}$$

It is easiest to estimate each of the factors in the right-hand side of the above equality separately. We bound the first factor, $|\delta^{D_1+2D_2m^*}F|_R$, through an application of the triangle inequality:

$$\begin{aligned} |\delta^{D_1+2D_2m^*}F|_R &= \left|\delta^{D_1+2D_2m^*}\sum_{k=0}^{D_1-1}\sum_{\ell=0}^{D_2-1} c_{k\ell}z^k\left(e^{\log\alpha z}\right)^{\ell}\right|_R \\ &\le \delta^{D_1+2D_2m^*}D_1D_2\max\{|c_{k\ell}|\}|z|_R^{D_1}\left|e^{\log\alpha z}\right|_R^{D_2} \\ &\le \delta^{D_1+2D_2m^*}D_1D_2c_2^{m^{3/2}\log m}R^{D_1}\left(e^{|\mathrm{Re}(\log\alpha)|R}\right)^{D_2} . \end{aligned}$$

In view of our choices of D_1 and D_2 and the fact that $R > m^* > m$, we see that the previous inequality implies

$$|\delta^{D_1+2D_2m^*}F|_R < c_7^{m^{3/2}\log R + m^{1/2}R}\ .$$

The second factor in the above equation satisfies

$$\prod_{a=1}^{m-1}\prod_{b=1}^{m-1}|(a^*-a)+(b^*-b)\beta| \le \Big(m^*(1+|\beta|)\Big)^{m^2}\ .$$

And, fixing $R = (4m)^{3/2}$ and taking m, and therefore m^*, so large that $R > m^*(1+|\beta|)$, it is straightforward to produce a *lower* bound for the denominator in the above equation: For $z = R$

$$\left|\prod_{a=1}^{m-1}\prod_{b=1}^{m-1}(z-(a+b\beta))\right|_R \ge \left|\prod_{a=1}^{m-1}\prod_{b=1}^{m-1}(R-(a+b\beta))\right| \ge \Big(R-m^*(1+|\beta|)\Big)^{m^2}\ .$$

We conclude that

$$\begin{aligned}\frac{\prod_{a=1}^{m-1}\prod_{b=1}^{m-1}|(a^*-a)+(b^*-b)\beta|}{\left|\prod_{a=1}^{m-1}\prod_{b=1}^{m-1}(z-(a+b\beta))\right|_R} &\le \left(\frac{m^*(1+|\beta|)}{R-m^*(1+|\beta|)}\right)^{m^2}\\ &\le c_8^{-3/2(m^*)^2\log(m^*)+m^2\log(m^*(1+|\beta|)}\\ &< c_9^{-1/2m^2\log m}.\end{aligned}$$

Estimating the secondary factors. As we indicated above, we estimate the secondary factors in N, the factors that involve one of the conjugates θ_i for $i = 2, 3, \ldots, d$, through the triangle inequality:

$$\begin{aligned}|A_1^* + A_2^*\theta_i + \cdots + A_d^*\theta_i^{d-1}| &\le d\max_{1\le j\le d}\{|A_j^*|\}\max\{1,|\theta_i|\}^{d-1}\\ &\le dc_6^{(m^*)^{3/2}\log m^*}\max\{1,|\theta_1|,|\theta_2|,\ldots,|\theta_d|\}^{d-1} \le c_{10}^{(m^*)^{3/2}\log m^*}\ ;\end{aligned}$$

Therefore

$$\begin{aligned}&\prod_{j=2}^{d}\left|A_1^* + A_2^*\theta_i + A_3^*\theta_i^2 + \cdots + A_d^*\theta_i^{d-1}\right|\\ &\qquad \le \left(c_{10}^{(m^*)^{3/2}\log m^*}\right)^{d-1} = \left(c_{10}^{d-1}\right)^{(m^*)^{3/2}\log m^*} \le c_{11}^{m^{3/2}\log m}.\end{aligned}$$

Therefore we are lead to the estimate for $|N|$:

$$|N| < c_{11}^{m^{3/2}\log m}c_9^{-1/2m^2\log m} \le c_{12}^{-1/2m^2\log m} < 1,$$

for m sufficiently large, which completes our proof.

How to obtain the nonzero value.

There are several ways to find an appropriate nonzero algebraic number. For historical accuracy we first consider Schneider's fairly complicated approach to this problem first.

Schneider's original method. To obtain his nonzero algebraic number Schneider considers several functions associated with the function $F(z)$. We will see that his reason for doing this is to find enough functions that, if they all vanish at the points under consideration, a certain Vandermonde matrix will vanish. We will see that this argument depends on an additional assumption about the algebraic nature of α. Specifically, Schneider points out that he may assume that α is not a root of unity. Indeed, if it is, then instead of considering the numbers $\alpha, \beta,$ and α^β at the very beginning of the proof one considers the numbers $\alpha^\beta, \beta^{-1},$ and α.

For notational simplicity in the argument below we retain the notation $D_1 = \sqrt{2d}\, m^{3/2}$ and $D_2 = \sqrt{2d}\, m^{1/2}$. Using this notation Schneider defined his associated functions as follows: for each $\sigma, 1 \leq \sigma \leq D_2$, let

$$F_\sigma(z) = \left[\prod_{1 \leq a \leq \sigma - 1, 1 \leq b \leq m} (z - (a + b\beta)) \right] F(z + \sigma - 1).$$

Notice that each $F_\sigma(z)$ vanishes at the prescribed zeros of $F(z)$, $a + b\beta, 1 \leq a, b \leq m$.

In order to understand the matrix Schneider introduces it it helpful to first rewrite the original auxiliary function as:

$$F(z) = P_{11}(z) + P_{12}(z)\alpha^z + P_{13}(z)\alpha^{2z} + \cdots + P_{1D_2}(z)\alpha^{(D_2-1)z}.$$

It is then an easy calculation to rewrite each of the function $F_\sigma(z), 1 \leq \sigma \leq D_2$, in terms of polynomials $P_{12}, \ldots, P_{1D_2}$ which may be easily described in terms of the polynomials $P_{1\tau}(z)$, above. Specifically, if for each pair σ, τ we put

$$P_{\sigma\tau}(z) = \prod_{1 \leq a \leq \sigma - 1, 1 \leq b \leq m} (z - (a + b\beta)) P_{1\tau}(z - \sigma - 1),$$

then we have

$$F_\sigma(z) = \sum_{\tau=1}^{D_2} \alpha^{(\sigma-1)(\tau-1)} P_{\sigma\tau}(z) \alpha^{(\tau-1)z}.$$

We note that the vanishing of all of these functions at the indicated points translates into have a certain matrix product equalling zero. Specifically for each z

$$\begin{pmatrix} P_{11}(z) & P_{12}(z) & \cdots & P_{1D_2}(z) \\ P_{21}(z) & \alpha P_{22}(z) & \cdots & \alpha^{D_2-1}P_{2D_2}(z) \\ P_{31}(z) & \alpha^2 P_{22}(z) & \cdots & \alpha^{2(D_2-1)}P_{3D_2}(z) \\ \vdots & \vdots & \vdots & \vdots \\ P_{D_21}(z) & \alpha^{D_2-1}P_{D_22}(z) & \cdots & \alpha^{(D_2-1)(D_2-1)}P_{D_2D_2}(z) \end{pmatrix} \times \begin{pmatrix} 1 \\ \alpha^z \\ \alpha^{2z} \\ \vdots \\ \alpha^{(D_2-1)z} \end{pmatrix} = \vec{0}.$$

By our application of Siegel's Lemma not all of the polynomials in the first row of the matrix are identically zero; we denote the nonzero polynomials in the first row by $P_{1\tau_1}(z), \ldots, P_{1\tau_r}(z)$ and consider the $r \times r$ matrix:

$$\begin{pmatrix} P_{1\tau_1}(z) & P_{1\tau_2}(z) & \cdots & P_{1\tau_r}(z) \\ \alpha^{(\tau_1-1)}P_{2\tau_1}(z) & \alpha^{(\tau_2-1)}P_{2\tau_2}(z) & \cdots & \alpha^{(\tau_r-1)}P_{2\tau_r}(z) \\ \alpha^{2(\tau_1-1)}P_{3\tau_1}(z) & \alpha^{2(\tau_2-1)}P_{3\tau_2}(z) & \cdots & \alpha^{2(\tau_r-1)}P_{3\tau_r}(z) \\ \vdots & \vdots & \vdots & \vdots \\ \alpha^{(r-1)(\tau_1-1)}P_{r\tau_1}(z) & \alpha^{(r-1)(\tau_2-1)}P_{r\tau_2}(z) & \cdots & \alpha^{(r-1)(\tau_r-1)}P_{r\tau_r}(z) \end{pmatrix}$$

Following Schneider we temporarily let

$$\Pi_\sigma(z) = \prod_{1 \le a \le \sigma-1, 1 \le b \le m} (z - (a + b\beta))$$

so we have

$$P_{\sigma\tau}(z) = \Pi_\sigma(z) P_{1\tau}(z - \sigma - 1),$$

Thus the above matrix may be represented by the product:

$$\begin{pmatrix} P_{1\tau_1}(z) & P_{1\tau_2}(z) & \cdots & P_{1\tau_r}(z) \\ \Pi_2(z)P_{1\tau_1}(z-1) & \Pi_2(z)P_{1\tau_2}(z-1) & \cdots & \Pi_2(z)P_{1\tau_r}(z-1) \\ \Pi_3(z)P_{1\tau_1}(z-2) & \Pi_3(z)P_{1\tau_2}(z-2) & \cdots & \Pi_2(z)P_{1\tau_r}(z-2) \\ \vdots & \vdots & \vdots & \vdots \\ \Pi_r(z)P_{1\tau_1}(z-r+1) & \Pi_r(z)P_{1\tau_2}(z-r+1) & \cdots & \Pi_r(z)P_{1\tau_r}(z-r+1) \end{pmatrix}$$
$$\times \begin{pmatrix} 1 & \alpha^{\tau_1-1} & \cdots & \alpha^{(r-1)(\tau_1-1)} \\ 1 & \alpha^{\tau_2-1} & \cdots & \alpha^{(r-1)(\tau_2-1)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \alpha^{\tau_r-1} & \cdots & \alpha^{(r-1)(\tau_r-1)} \end{pmatrix}$$

We denote the second matrix above by W and note that its determinant is Vandermonde. Then if we let $a_j x^{g_j}$ denote the leading coefficient of $P_{1\tau_j}(z)$ the determinant of the above product may be written as

$$\begin{aligned} & D(z) \\ &= \big(\Pi_1(z)\cdots\Pi_k(z)\big)\big(a_1 \cdots a_r z^{g_1+\cdots+g_r} \times |W| + \text{ lower degree terms } \times |W|\big). \end{aligned}$$

If $D(z)$ vanishes identically then the coefficient of each power of z must equal zero. But the leading coefficient of $D(z)$ equals zero only if $|W| = 0$, which would imply that α is a root of unity, contrary to our earlier assumption.

Schneider next shows that the function $D(z)$ is a polynomial in z, with coefficients involving powers of α; the degree of $D(z)$ may be shown to be less than $12m^2$. Thus there exists a pair $a^* + b^*\beta, 1 \leq a^*, b^* < 4m$ so that $D(a^* + b^*\beta) \neq 0$. This means that none of the rows of the above matrix can vanish at $a^* + b^*\beta$ and, looking at the first row, we deduce that $F(a^* + b^*\beta) \neq 0$.

Alternate ways to obtain the nonzero value. We saw above that Schneider used a subtle argument, based on the nonvanishing of a Vandermonde determinant, to obtain a point $a^* + b^*\beta$ which produced a nonzero algebraic number $F(a^* + b^*\beta)$ that eventually lead to a positive integer less than 1. It is obvious that obtaining such a nonzero algebraic number was central to both Gelfond's and Schneider's methods. Perhaps not unexpectedly, finding alternate approaches to finding a nonzero value for large classes of analytic functions became an important area of research in transcendental number theory in the second half of the twentieth century. We conclude this chapter with two other approaches to obtaining the all-important nonzero value for an exponential function.

Using the growth of the function to find a nonzero value. The first approach is based on considering the following question: How many zeros can a nonzero entire function have in a sequence of increasing large discs? The motivation for this question is based on the following observation, which we will quantify: If an entire function has "too many" zeros in an increasing sequence of discs, then the function must be identically zero. The quantitative version of this qualitative observation is given below. We first make an important definition.

Given a function $F(z)$ and a positive real number R, we define the set

$$Z(F, R) = \{z \in \mathbf{C} : F(z) = 0, \text{ with } |z| \leq R \} ,$$

with the understanding that if $z_0 \in Z(F, R)$ is a zero of $F(z)$ with multiplicity m, then z_0 appears m times in the set $Z(F, R)$. Finally, we denote the cardinality of the set $Z(F, R)$ by $\mathrm{card}(Z(F, R))$.

Theorem (Order of growth and zeros). *Let $F(z)$ be an entire function and suppose that there exists a real number κ such that for all sufficiently large R,*

$$\max_{|z| \leq R} \{|F(z)|\} = |F|_R \leq e^{R^\kappa} .$$

If there exists an $\epsilon > 0$ and an unbounded sequence of real numbers $R_1, R_2, \ldots$ such that

$$R_n^{\kappa+\epsilon} < \mathrm{card}(Z(F, R_n)) ,$$

for all $n = 1, 2, \ldots$, then $f(z)$ is the identically zero function.

Proof . Given any $z_0 \in \mathbf{C}$, there exists an integer n_0 such that for all $n \geq n_0$, we have $|z_0| < R_n$. For each $n \geq n_0$, we define

$$G_n(z) = \frac{F(z)}{\prod_{\omega \in Z(F,R_n)}(z-\omega)} \ .$$

Thus, given that the zeros in $Z(F, R_n)$ appear with the appropriate multiplicity, we see that $G_n(z)$ is an entire function. Hence by the Maximum Modulus Principle, and our hypotheses, we conclude that for all sufficiently large R_n,

$$\begin{aligned}
|F(z_0)| &\leq |G_n|_{5R_n} \prod_{\omega \in Z(F,R_n)} |z_0 - \omega| \\
&= \frac{|F|_{5R_n}}{\left|\prod_{\omega \in Z(F,R_n)}(z-\omega)\right|_{5R_n}} \prod_{\omega \in Z(F,R_n)} |z_0 - \omega| \\
&\leq \frac{e^{(5R_n)^\kappa}}{(4R_n)^{\mathrm{card}(Z(F,R_n))}} (2R_n)^{\mathrm{card}(Z(F,R_n))} \\
&\leq e^{(5R_n)^\kappa} \left(\frac{1}{2}\right)^{\mathrm{card}(Z(F,R_n))} \\
&< e^{(5R_n)^\kappa} \left(\frac{1}{2}\right)^{R_n^{\kappa+\epsilon}} \\
&= e^{(5R_n)^\kappa} e^{-(\log 2)R_n^{\kappa+\epsilon}} \\
&= e^{(5^\kappa - (\log 2)R_n^\epsilon)R_n^\kappa}.
\end{aligned}$$

Since for all sufficiently large n, we have $5^\kappa < (\log 2)R_n^\epsilon$, we conclude that if we let $n \to \infty$, then $|F(z_0)| = 0$, and hence $F(z_0) = 0$ for all $z_0 \in \mathbf{C}$, which establishes our result.

Let's see how the above theorem can be applied to simplify Schneider's conclusion. In order to estimate the number of zeros of the function $F(z)$ in the Main Proposition,

$$F(z) = P(z, e^{\log \alpha z}) = \sum_{k=0}^{D_1-1} \sum_{\ell=0}^{D_2-1} c_{k\ell} z^k e^{\ell \log \alpha z}$$

we need to find a value for the exponent κ that fulfills the hypothesis of the above theorem.

Claim. For any choice of $\kappa > 1$ for all sufficiently large R,

$$|F|_R \leq e^{R^\kappa} \ .$$

In order to establish this claim we note that it follows from the Maximum Modulus Principle and the triangle inequality that

$$\begin{aligned} |F|_R &= \max_{|\zeta|=R}\{D_1 D_2 \max\{|c_{k\ell|}|\}|\zeta|^{D_1} e^{D_2|\log\alpha||\zeta|}\} \\ &\le D_1 D_2 \max\{|c_{k\ell|}|\} R^{D_1} e^{D_2|\log\alpha|R} \\ &\le e^{R^\kappa} \text{ for any } \kappa > 1, \end{aligned}$$

where these last inequalities follow since if we take R sufficiently large, since the quantities $D_1, D_2,$ and $\max\{|c_{k\ell|}|\}$ do not depend on R. This establishes the claim.

Thus, since we have already established that $F(z)$ is not identically zero (since the functions z and $e^{\log\alpha z}$ are algebraically independent and the coefficients $c_{k\ell}$ are not all equal to 0), we see that for any choice of $\epsilon > 0$, for all sufficiently large R, $F(z)$ cannot have more than $R^{\kappa+\epsilon}$ zeros in $Z(F,R)$.

More concretely, taking $\kappa = \frac{5}{4}$ and $\epsilon = \frac{1}{4}$, we see that for all sufficiently large R,

$$\mathrm{card}(Z(F,R)) \le R^{3/2} \ .$$

Conclusion. Using the above estimate for $\mathrm{card}(Z(F,R))$ for sufficiently large R, it is possible to deduce that there exists an integer $M \ge m$ such that

$$F(a+b\beta) = 0 \text{ for all } 1 \le a < M \textit{ and } 1 \le b < M \ ,$$

while there exists a pair a^*, b^* with $1 \le a^* \le M, 1 \le b^* \le M$, such that $F(a^* + b^*\beta) \ne 0$. (*Hint:* Assume that $F(a+b\beta) = 0$ for all positive integers a and b. Now show that for all sufficiently large R, a lower bound for the number of complex numbers of the form $z = a + b\beta$, where a and b are positive integers, that satisfy $|z| \le R$ is proportional to the area of the disk of radius R. Conclude that this assumption implies that $F(z)$ is identically zero, which is a contradiction.)

A More Modern Approach to Obtaining a Nonzero Value: A Zeros Estimate

This approach is based on providing a count of the total number of zeros a so-called *exponential polynomial* can have. What is surprising about this approach is that, at least in the real case, the proof requires no ideas beyond basic calculus. (This proposition, due to Polya, was used by Gelfond in 1962 when he provided a simpler proof of his theorem in case both α and β are real.)

Proposition. *Let $P_1(z), P_2(z), \ldots, P_k(z)$ be polynomials with real coefficients and degrees $d_1, d_2, \ldots, d_k$, respectively. Suppose $\omega_1, \omega_2, \ldots, \omega_k$ are distinct real numbers. Then the function*

$$F(z) = P_1(z)e^{\omega_1 z} + P_2(z)e^{\omega_2 z} + \cdots + P_k(z)e^{\omega_k z}$$

has at most

$$d_1 + d_2 + \cdots + d_k + k - 1$$

real zeros.

Proof. We note for later use that, after multiplying $F(z)$ by $e^{-\omega_k z}$ we may assume that $\omega_k = 0$.

The proof is by induction on $n = d_1 + d_2 + \cdots + d_k + k$. If $n = 1$ then $k = 1$ and $d_1 = 0$. Thus $F(z) = a_1 e^{\omega_1 z} = a_1$, by our simplifying assumption above. Since $a_1 \neq 0$ $F(z)$ has no zeros.

We now take $m \geq 2$, assume the result has been established for all functions with $n = d_1 + d_2 + \cdots + d_k + k < m$, and let $F(z)$ be a function as above with $d_1 + d_2 + \cdots + d_k + k = m$. Let N denote the number of real zeros of $F(z)$. The trick is to apply Rolle's Theorem, by which we know that the number of zeros of

$$\begin{aligned} F'(z) &= \omega_1 P_1(z) e^{\omega_1 z} + P_1'(z) e^{\omega_1 z} + \omega_2 P_2(z) e^{\omega_2 z} + P_2'(z) e^{\omega_2 z} + \cdots + P_k'(z) \\ &= (\omega_1 P_1(z) + P_1'(z))\, e^{\omega_1 z} + (\omega_2 P_2(z) + P_2'(z))\, e^{\omega_2 z} + \cdots + P_k'(z), \end{aligned}$$

is at least $N - 1$.

Notice that in the above representation of $F'(z)$ we have for $j = 1, \ldots, k - 1, \deg(\omega_j P_1(z) + P_j'(z)) \leq d_j$. However the degree of the coefficient of the term e^0 is one less than the degree of the coefficient of e^0 in $F(z)$. Therefore we may apply the induction hypothesis to conclude that

$$N - 1 \leq d_1 + d_2 + \cdots + d_k + k - 2,$$

from which the proposition follows.

Exercises.

1. For an arbitrary algebraic number α of degree $\deg(\alpha) = d$, let

$$P(x) = c_d x^d + c_{d-1} x^{d-1} + \cdots + c_0 \in \mathbf{Z}[\mathbf{x}]$$

denote the minimal polynomial for α; thus $\gcd(c_0, c_1, \ldots, c_d) = 1$. We define the *height of* α, denoted by $H(\alpha)$, to be the height of its minimal polynomial. That is,

$$H(\alpha) = H(P) = \max\{|c_0|, |c_1|, \ldots, |c_d|\}\ .$$

Then there exist rational numbers $c_{0,n}, c_{1,n}, \ldots,\ c_{d-1,n}$ such that

$$\alpha^n = c_{0,n} + c_{1,n}\alpha + c_{2,n}\alpha^2 + \cdots + c_{d-1,n}\alpha^{d-1}\ .$$

Moreover,

$$\max_{0 \leq j \leq d-1} \{|c_{j,n}|\} \leq (1 + H(\alpha))^{n+1-d}\ ,$$

and each rational number $c_{j,n}$ can be expressed as a fraction having a denominator equal to c_d^{n+1-d}.

2. Suppose $\beta_1, \beta_2, \ldots, \beta_L$ are elements of $\mathbf{Q}(\theta)$, where θ is an algebraic integer of degree d and of height $H(\theta)$. If for each $l = 1, 2, \ldots, L$,

$$\beta_l = r_{l1} + r_{l2}\theta + \cdots + r_{ld}\theta^{d-1} \ ,$$

where each r_{lj} is a rational number satisfying $|r_{lj}| \le B_l$ for some bound B_l, then

$$\beta_1\beta_2 \cdots \beta_L = r_1 + r_2\theta + \cdots + r_d\theta^{d-1} \ ,$$

with rational coefficients r_j satisfying

$$\max_{1\le j\le d}\{|r_j|\} \le d^L B_1 B_2 \cdots B_L \Big(2H(\theta)\Big)^{dL}$$

Moreover, if $\mathrm{den}(\beta_l)$ denotes the least common multiple of the denominators of the rational coefficients $r_{l1}, r_{l2}, \ldots, r_{ld}$, then each rational number r_j has a denominator of the form

$$\mathrm{den}(\beta_1)\mathrm{den}(\beta_2)\cdots\mathrm{den}(\beta_L) \ .$$

3. Convince yourself that the details in the proof of the zeros estimate concerning order of growth are correct.

4. Using the nonzero value from the order of growth estimate to obtain the final contradiction in Schneider's proof.

5. In Schneider's proof assume that α and β are *real* numbers. Apply the last proposition of this chapter to show that there exists a constant c^* so that c^*m satisfies: There exists a pair a^*, b^*, with $\max\{a^*, b^*\} = m^*$, such that $F(a^* + b^*\beta) \ne 0$ but for all a, b with $1 \le a, b < c^*m, F(a + b\beta) = 0$.