# Chapter Seven
# Variants and Alternate Points of View

In previous chapters we considered many of the ideas that came to dominate transcendence theory in the twentieth century. There are a few more that we will explore in the remaining pages. These involve how authors sought to generalize the Gelfond-Schneider result, a method that leads to a nonzero, algebraic value without explicitly requiring an *auxiliary* function, and the Gelfond-Schneider result in other settings.

The first of these involves both modifying the initial assumption that $\alpha$ and $\beta$ be algebraic and providing quantitative versions of the Gelfond-Schneider Theorem. In this lecture we first survey the mid-twentieth century generalizations of this theorem and we then consider an idea attributed to Michel Laurent that avoids explicitly displaying the auxiliary function (and leads to better results, especially with regards to constants, than had been obtained with explicit functions).

Generalizations to the Gelfond-Schneider Theorem began to appear almost immediately after the publication of Gelfond's proof in 1934 and of Schneider's proof in 1935. In 1935 G. Ricci used Gefond's approach and introduced a special type of Liouville number into the statement of the Gelfond-Schneider Theorem. Ricci established several theorems with rather complicated statements. Ricci's most easily stated theorem contains the following result.

**Theorem.** Suppose $\alpha$ are $\beta$ are algebraic numbers where $\alpha \neq 0, 1$ and $\beta$ is irrational. Suppose further that $\kappa$ is an irrational number such that for some $\epsilon > 0$

$$\left|\kappa - \frac{p}{q}\right| < e^{-(\log q)^{2+\epsilon}} \tag{1}$$

has infinitely many solutions $p \in \mathbf{Z}, q \in \mathbf{N}$. Then $(\kappa\alpha)^{\beta}$ is transcendental.
This is Teorema VI, part 2 of [Ri].

Note: The number $\kappa$ in the statement of Ricci's Theorem is transcendental by Liouville's Theorem, from Chapter One. An example of such a number $\kappa$ is a decimal having appropriately increasing sequences of zeros. Moreover, the number $\kappa\alpha$ in the conclusion of this theorem is also transcendental.

In 1937 P. Franklin published a different sort of generalization. Instead of using Liouville numbers, which are very well approximated a sequence of rational numbers, Franklin considered numbers that well-approximated by algebraic numbers from a fixed number field. In order to appreciate how far the field has advanced, let's look the full statement of one of Franklin's Theorems.

**Theorem.** Let $\{\alpha_i\}, \{\beta_i\},$ and $\{\eta_i\}$ be three sequences of irrational numbers in a fixed number field K, where the conjugates of all of the elements of these sequences are uniformly bounded. Let $\delta_i$ be a sequence of integers, becoming infinite, such that $\delta_i\alpha_i, \delta_i\beta_i,$ and $\delta_i\eta_i$ are algebraic integers. Suppose

$$a = \lim_{n\to\infty} \alpha_i, b = \lim_{n\to\infty} \beta_i \text{ and } a^b = \lim_{n\to\infty} \eta_i, \tag{2}$$

where $a \neq 0, 1, b \neq 0, 1$. For each $i$ let $\delta_i$ be a denominator for each of the numbers $a_i$, and $b_i$. Then if for some $\kappa > 6$ we have

$$|a - \alpha_i| + |b - \beta_i| + |a^b - \eta_i| < \delta_i^{-(\log \delta_i)^\kappa}, \tag{3}$$

then

$$\kappa > 6.$$

This is Theorem I, page 162, of [Fr], with $A = a^b, B = a,$ and $H = b$.

In another direction, within a few years of the solution to the $\alpha^\beta$ portion of Hilbert's seventh problem, authors began to provide quantitative interpretations of the Gelfond-Schneider Theorem. These quantitative results can take one of two forms depending on how you view the statement: For algebraic numbers $\alpha$ and $\beta$, with $\alpha \neq 0, 1$ and $\beta$ irrational, $\alpha^\beta$ is transcendental. One is to conclude that for any nonzero integral polynomial $P(x)$, $P(\alpha^\beta) \neq 0$. Another is to conclude that for any algebraic number $\gamma, \alpha^\beta - \gamma \neq 0$. It is these two, related, statements that were first given quantitative versions.

These early results, as with almost all subsequent such results, had very explicit, so fairly complicated, statements. We state two of them, due to Gelfond, here. But before we do we need to introduce a new concept, the height of a polynomial (and the height of an algebraic number). For any polynomial $P(x)$ the *height* of $P$, denoted $H(P)$ equals the maximum absolute value of its coefficients. The height of an algebraic number $\alpha$, denoted $H(\alpha)$, is the height of its minimal integral polynomial. Of course the degree of an algebraic number is the degree of its minimal polynomial.

**Theorem** (Gelfond, 1935). Suppose $\alpha$ and $\beta$ are algebraic numbers with $\alpha \neq 0, 1$ and $\beta$ irrational.

1. Take $d \in \mathbf{N}$ and $\epsilon > 0$. There exists a constant $c(\alpha, \beta, d, \epsilon)$ such that for any algebraic number $\gamma$ with $deg(\gamma) \leq d$ and $H(\gamma) > c$,

$$|\alpha^\beta - \gamma| > H(\gamma)^{-(\log \log H(\gamma))^{5+\epsilon}}.$$

2. Suppose $P(x) \in \mathbf{Z}[x]$ has degree $d$ and height $H(P)$. Then there exists a constant $c$ such that

$$|P(\alpha^\beta)| > e^{-cd^2(d+\log H(P)) \log^2(d+\log H(P)+1) \log^{-3}(d+1)}.$$

These two admittedly difficult to grasp results were greatly improved throughout the twentieth century. We do not pursue this history here; instead we look at two fairly recent results.

### Modern Generalizations of the Gelfond-Schneider Theorem

One thing shared by the two theorems we are going to look at is their method of proof–instead of relying on Siegels' Lemma to establish the existence of an advantageous function, they work directly with the matrix that would underlie an application of that lemma. This idea, of looking directly at the matrix instead of deducing from the matrix the existence of a function with certain properties, began with the work of Michel Laurent in the early 1990s.

This first result was formulated, and established, by Michel Waldschmidt. While it is in the spirit of Franklin's result, in that it simultaneously says something about the arithmetic nature of three values related to the Gelfond-Schneider Theorem, we will see that its proof is straightforward. Alas, in order to avoid stating a very awkward version of this and the other results of this section, we need to introduce a bit of standard terminology.

**Theorem.** Let $\alpha$ be a positive real number, $\alpha \neq 1$, and $\beta$ an irrational real number. Let $\alpha^\beta = e^{\beta \log \alpha}$, where $\log \alpha$ is the real value of the logarithm of $\alpha$. Then for any sufficiently large rational integer $N$ there exists a polynomial $P \in \mathbf{Z}[X_1^{\pm}, X_2^{\pm}, Y]$ satisfying

$$\deg P + \log Ht(P) \leq 5N^{14} \log N,$$

and

$$0 < |P(\alpha, \alpha^\beta, \beta)| < e^{-\frac{1}{3}N^{16}}. \tag{4}$$

Before we examine the proof of this theorem, and its use of a determinant in place of a function constructed through an application of the pigeonhole principle, let's look at one particularly appealing corollary that greatly generalizes the Gelfond-Schneider Theorem. consequences.

**Corollary** Under the hypothesis of the above theorem not all of $\alpha, \beta,$ and $\alpha^\beta$ can be algebraic.

**Proof.** Suppose each of the numbers $\alpha, \beta,$ and $\alpha^\beta$ is algebraic. Let $\alpha_1(= \alpha), \ldots, \alpha_{d_1}$ denote the conjugates of $\alpha$; $\beta_1(= \beta), \ldots, \beta_{d_2}$ denote the conjugates of $\beta$, and $\gamma_1(+\alpha^\beta), \ldots, \gamma_{d_3}$ denote the conjugates of $\alpha^\beta$. For simplicity we assume $P$ does not involve negative powers of the variables (if it does we can multiply through by the variable to the appropriate power. Then

$$M = |P(\alpha, \alpha^\beta, \beta)| \prod_{(i,j,k)\neq(1,1,1)} |P(\alpha_i, \gamma_j, \beta_k)|$$

is a nonzero integer.

Using the estimate of the theorem for $|P(\alpha, \alpha^\beta, \beta)|$, and estimating each of the other terms by the triangle inequality, shows that by taking $N$ sufficiently large the integer $M$ satisfies: $0 < |M| < 1$.

*The proof of Waldschmidt's Theorem.* The proof of the above theorem does not look like any of the other proofs we have considered but we will see that it has the same essential components (as the following outline indicates).

**Outline of the proof.**

**0.** For simplicity, although risking confusing the reader, we let

$$D_1 = N^6 - 1,\ D_2 = \frac{1}{2}(N^2 - 1),\ \text{and } K = \frac{1}{2}(N^4 - 1),$$

and we restrict $N$ to be odd so that each of the above parameters is an integer.

**1.** Instead of looking at an auxiliary function of the form

$$F(z) = \sum_{m=0}^{D_1} \sum_{n=-D_2}^{D_2} a_{mn} z^m \alpha^{nz}$$

at points $k_1 + k_2\beta, -K \leq k_1, k_2 \leq K$, for some parameters $D_1, D_2,$ and $K$ we just consider the collection of functions

$$\phi_{mn} = z^m \alpha^{nz}, 0 \leq m \leq D_1, -D_2 \leq n \leq D_2.$$

We can put any ordering we want on this collection of functions; for clarity we order them lexicographically:

$$\phi_{0,-D_2}, \phi_{0,-D_2+1}, \ldots, \phi_{0,D_2}, \phi_{1,-D_2}, \ldots, \phi_{1,D_2}, \ldots \phi_{D_1,-D_2}, \ldots, \phi_{D_1,D_2}.$$

which we respectively label $\phi_1, \ldots, \phi_L,\ L = (D_1 + 1)(2D_2 + 1)$.

We evaluate each of these functions at the points $\xi_{k_1,k_2} = k_1 + k_2\beta, -K \leq k_1, k_2 \leq K$; we also order these points lexicographically and then label them according to this ordering: $\zeta_1, \ldots, \zeta_{2K+1}$.

The remainder of the proof works directly with the matrix which consists of the functions $\phi_{mn}$ evaluated at the points $k_1 + k_2\beta$. The columns of the matrix are indexed by the ordering of the functions $\phi_{mn}$ and the rows are indexed by the ordering of the points $k_1 + k_2\beta$. Since we want this to be a square matrix we take

$$L = (D_1 + 1)(2D_2 + 1) = (2K + 1)^2,$$

and consider the matrix:

$$\big(\phi_\lambda(\zeta_\mu)\big)_{1 \leq \lambda, \mu \leq L}. \tag{5}$$

(The polynomial $P$ in the conclusion of the theorem is the determinant of the matrix above with $X_1$ replacing $\alpha$, $X_2$ replacing $\alpha^\beta$, and $Y$ replacing $\beta$.)

**2.** We will denote the determinant of this matrix by $\Delta$. The zeros estimate at the end of Chapter 5 may be applied to conclude that $\Delta \neq 0$.

**3.** The degree and height of the determinant of the matrix with $X_1$ replacing $\alpha_1$, $X_2$ replacing $\alpha^\beta$, and $Y$ replacing $\beta$ are computed directly from representing this determinant as a sum of products of its entries.

**4.** The nonzero value from the first step is then estimated from above through an application of the Maximum Modulus Principle.

**Details of the proof.**

*Application of the Zeros Estimate.* In order to apply the zeros estimate from the end of Chapter 5 to show that $\Delta$ does not vanish we need to show that if it did vanish we would have a nonzero exponential polynomial with too many real zeros. If $\Delta = 0$ then its columns are linearly dependent. Using our ordering we can explicitly represent this linear dependency between the columns as:

$$\sum_{\lambda=1}^{L} A_\lambda \phi_\lambda(z)\Big|_{z=\zeta_\mu} = 0,\ 1 \le \mu \le 2K+1. \tag{6}$$

This dependency may be translated into our previous function notation:

$$F(z) = \sum_{m=0}^{D_1} \sum_{n=-D_2}^{D_2} a_{mn} z^m \alpha^{nz}\Big|_{(k_1+k_2\beta)} = 0 \tag{7}$$

where the coefficients $a_{mn}$ are not all zero. Thus, if $\Delta = 0$ the above function, $F(z)$, vanishes at the $L$ distinct points $k_1 + k_2\beta$.

The application of the zeros estimate is clearer if we rewrite $F(z)$ as

$$F(z) = \sum_{n=-D_2}^{D_2} \left( \sum_{m=0}^{D_1} a_{mn} z^m \right) e^{\omega_n z},$$

where $\omega_n = n \log \alpha$. Then, according to the zeros estimate,$F(z)$ can have at most

$$\underbrace{D_1 + \cdots + D_1}_{(2D_2+1)\text{terms}} + (2D_2+1) - 1 = L - 1,$$

zeros. Thus not all of the values in (8) can equal zero, so not such dependency (7) can hold. It follows that $\Delta \neq 0$.

*Estimating the Degree and Height of $\Delta$:* We introduce *monomials* $P_{mn}^{k_1,k_2} = (k_1 + k_2Y)^m (X_1^{k_1} X_2^{k_2})^n$, where we italicized monomials since they can have negative degrees, so that

$$P_{mn}^{k_1,k_2}(\alpha, \alpha^\beta, \beta) = \phi_{mn}(k_1 + k_2\beta).$$

If we use the lexicographical ordering on the subscripts of these polynomials to index the columns of a matrix, and use the lexicographical ordering on the superscripts of these polynomials to index the rows of a matrix, then the matrix we considered above,

$$\big(\phi_\lambda(\zeta_\mu)\big)_{1 \le \lambda, \mu \le L}.$$

is the same as the matrix

$$\big(P_{mn}^{k_1,k_2}\big) \tag{8}$$

evaluated at $X_1 = \alpha, X_2 = \alpha^\beta$, and $Y = \beta$. Notice that we have the following easy estimates:

$$\deg P_{mn}^{k_1,k_2} \leq (D_1 + 2D_2K) \text{ and } Ht(P_{mn}^{k_1,k_2}) \leq (2K)^{D_1}. \tag{9}$$

The determinant of the matrix (7) is a polynomial expression in $X_1^\pm, X_2^\pm, Y$, which we denote by $P$. In order to estimate the degree and height of $P$ we note that it is a signed sum of terms each of which is a product of $L$ polynomials, one from each row and each column of the matrix (9). From this observation we have the easy estimate

$$\deg P \leq L(D_1 + 2D_2K).$$

Estimating the height of $P$ is a more complicated matter. Indeed it is easier to do if we introduce yet-another new concept. For a polynomial $Q$ we define the *length* of $Q$, denoted $L(Q)$, to be the sum of the absolute values of its coefficients. The reason the length of a polynomial is such a useful concept is because the following simple relationships hold: For any polynomials $P$ and $Q$,

$$L(P+Q) \leq L(P) + L(Q) \text{ and } L(PQ) \leq L(P)L(Q).$$

Notice that each of the monomials $P_{mn}^{k_1,k_2}$ has length at most $(2K)^{D_1}$. So from the above characterization of the determinant, using the above two identities about the lengths of the sum or products of two polynomials, we have

$$L(P) \leq L!(\max\{L(P_{mn}^{k_1,k_2})\})^L \leq L!(2K)^{LD_1}.$$

But, clearly $Ht(P) \leq L(P)$.

*An Upper Bound for* $\Delta$*:* So far we have established the lower bound $|\Delta| = |P(\alpha, \alpha^\beta, \beta)| > 0$. The upper bound is easier; it can be deduced from the following lemma.

**Lemma.** Let $R > r$ be two positive real numbers and suppose that $f_1(z), f_2(z), \ldots, f_L(z)$ are functions analytic in a set containing the disc $D = \{z : |z| \leq R\}$. Suppose that $\zeta_1, \ldots, \zeta_L$ all have absolute value at most $r$. Then the determinant

$$\Delta = \begin{pmatrix} f_1(\zeta_1) & \cdots & f_L(\zeta_1) \\ \vdots & \ddots & \vdots \\ f_1(\zeta_L) & \cdots & f_L(\zeta_L) \end{pmatrix}$$

satisfies

$$|\Delta| \leq \left(\frac{R}{r}\right)^{-L(L-1)/2} L! \prod_{\lambda=1}^{L} \max_{|\zeta|=R}\{|f_\lambda(\zeta)|\}.$$

*Sketch of proof.* The idea of this proof is to introduce a new variable $z$ and consider the function:

$$h(z) = \det\big(f_\lambda(\zeta_j z)\big)$$

and show that $h(z)$ has a zero at $z = 0$ to order at least $L(L-1)/2$. The key to this is to replace each of the functions $f_\lambda(\zeta_j z)$ with its Taylor series expansion at the origin and apply the multi-linearity of the determinant. This allows us to reduce the problem to the case of functions $f_\lambda(z) = z^{n_\lambda}, 1 \le \lambda \le L$, where each $n_\lambda$ is a non-negative integer. In this simple case we have

$$h(z) = z^{n_1+n_2+\cdots+n_L} \det\left(\zeta_\lambda^{n_\lambda}\right).$$

If $h(z)$ is not identically zero then the Vandermonde determinant of the matrix $\left(\zeta_\lambda^{n_\lambda}\right)$ is nonzero. Thus the non-negative integers $n_1, \ldots, n_L$ are pairwise distinct. Thus the sum $n_1 + \cdots + n_L$ is at least $0 + 1 + \cdots + (L-1)$ which equals $L(L-1)/2$. Implying that the order of vanishing of $h(z)$ at the origin is at least $L(L-1)/2$.

We now use this zero at $z = 0$ of order at least $L(L-1)/2$ to obtain the desired upper bound. The function

$$D(z) = \frac{h(z)}{z^{-L(L-1)/2}}$$

is analytic in the disc $|z| \le R$, and since $r < R$ we have $|D|_r \le |D|_R$. By the Maximum Modulus Principle we also have

$$|D|_r = r^{-L(L-1)/2}|h|_r \;\text{ and }\; |D|_R = R^{-L(L-1)/2}|h|_R.$$

Thus: $|h|_r \le \left(\frac{R}{r}\right)^{-L(L-1)/2}|h|_R$. If we now imagine using this inequality with 1 replacing $r$ and $R/r$ replacing $R$. Expanding the determinant we get $L!$ terms each being plus or minus a sum of elements one taken from each column and one taken from each row. We use the Maximum Modulus Principle to bound each of these terms and obtain the result of the lemma.

We apply this lemma with $r = K(1 + |\beta|)$ and $R = er$ to obtain the upper bound on $|\Delta|$ of the Proposition.

This concludes the proof of the theorem.

In our statement of the above result we glossed over many of the subtleties in Waldschmidt's formulation. By taking great care with each estimate Waldschmidt established the more widely applicable result we state below.

**Theorem.** Let $\alpha_1 \ne 0, 1$ be a positive real number, and let $\beta$ be an irrational real number. Put $\alpha_2 = \alpha_1^\beta = e^{\beta \log \alpha_1}$, where $\log \alpha_1$ is the real value of the logarithm of $\alpha_1$. Then for any rational integers $L, T_0, T_1$, and, $S$, and for any real number $E$, which satisfy:

$$T_0 \ge 2,\;\; T_1 \ge 2,\;\; S \ge 3,\;\; E \ge e,\;\; \text{and } L = (T_0+1)(2T_1+1) = (2S+1)^2,$$

there exists a polynomial $P \in \mathbf{Z}[X_1^{\pm}, X_2^{\pm}, Y]$ with

$$degP \le L(T_0 + 2T_1 S),\;\; Ht(P) \le L!(2S)^{LT_0}$$

and

$$0 < |P(\alpha_1, \alpha_2, \beta)| \le E^{-L^2/2}(SE)^{c_1 T_0 L} e^{c_1 T_1 SEL}$$

where $c_1 = c_1(\alpha_1, \beta) = (2 + |\beta|)(1 + |\log \alpha_1|)$.

The proof of this Theorem follows the same outline as the one we gave for the earlier Proposition. Indeed that Proposition can be deduced from this Theorem by specializing the parameters.

The method used to establish Waldschmidt's result has been fruitfully developed by several mathematicians to obtain some very precise results. We conclude this lecture with a single such result which has for its corollaries generalizations of several of the theorems we have seen. In order to state this result in its most general form need the notion of the Weil height of an algebraic number $\alpha$. To do this we assume $\alpha$ is of degree $d$ and let $\alpha_1, \alpha_2, \ldots, \alpha_d$ denote its conjugates. Then the Weil height of $\alpha$, denoted $h(\alpha)$ is defined as:

$$h(\alpha) = \frac{1}{d} \sum_{k=1}^{d} \log \max\{1, |\alpha_k|\}.$$

**Theorem (Nesterenko-Waldschmidt, 1996).** Let $\alpha$, and,$\beta$ be algebraic numbers. Put $K = \mathbf{Q}(\alpha, \beta)$ and let $D = [K : \mathbf{Q}]$. Let $A, B$ and,$E$ be positive real numbers satisfying:

$$A \ge exp(\max\{h(\alpha), D^{-1}\}, \quad B \ge h(\beta), \quad E \ge e. \tag{10}$$

Then for any nonzero complex number $\theta$,

$$\begin{aligned} &|e^\theta - \alpha| + |\theta - \beta| \\ &\quad \ge exp\Big( -211D\big(\log B + \log\log A + 4\log D + 2\log(E\max\{|\theta|, 1\}) + 10\big) \\ &\quad \times \big(D\log A + 2E|\theta| + 6\log E\big) \times \big(3.3D\log(D+2) + \log E\big) \times \big(\log E\big)^{-2}\Big). \end{aligned} \tag{11}$$

**Proof.** This is Theorem 1 (Main Theorem) of [Ne-Wa 1996].

Remark: This theorem in a precise quantitative version of the Hermite-Lindemann Theorem. Instead of saying simply that for any nonzero complex number $\theta$ at least one of the two numbers $\theta$ and $e^\theta$ is transcendental, it says a bit more. It tells us how well-approximated by algebraic numbers these two values can be.

**Corollary 1.** (A quantitative version of Lindemann's Theorem) For any real algebraic number $x$ with $\deg(x) = d$ and $L(x) = L \ge 3$,

$$|\pi - x| > \exp(-1.2 \times 10^6 d(\log L + d\log d)(1 + \log d)).$$

**Proof** Put $\theta = \pi i, \alpha = -1, \beta = ix, e = e^2, \log A = D^{-1},$ and $\log B = h(x) = h(\beta)$. Given these choices we have $D \leq 2d$.

**Corollary 2.** (A quantitative version of the Hermite-Lindemann Theorem) Suppose $\alpha$ and $\beta$ are algebraic numbers, with $\beta \neq 0$. Let $D = [\mathbf{Q}(\alpha, \beta) : \mathbf{Q}], E \geq e, A > 0$ with $\log A \geq \max\{h(\alpha), D^{-1} \log E, D{-}1|\beta|E\}$. Then

$$|e^\beta - \alpha| \geq \exp(-105,500 D^2 \log A(h(\beta) + \log \max\{1, \log A\} + \log D + \log E)(D \log D + \log E)(\log E)^{-2}).$$

**Exercises**

1. Derive the above lower bound for $|e^\beta - \alpha|$ in Corollary 2 from the general theorem.

2. Using any of the results of this chapter derive a quantitative version of Hermite's Theorem. Specifically, suppose $\alpha$ is an algebraic numbers. Obtain a lower bound for $|e - \alpha|$.

3. Suppose $P$ and $Q$ are polynomials in one variable. Establish the following inequalities involving their lengths $L(P)$ and $L(Q)$ :

A. $L(P) + L(Q) \leq L(P) + L(Q)$

B. $L(P)L(Q) \leq L(P)L(Q)$