A version of Baker's theorem on linear forms in logarithms

Adam J Harper

1st December 2010

Abstract

We reproduce a proof of a fairly weak version of Baker's theorem on linear forms in the logarithms of algebraic numbers. We try to motivate the argument by analogy with a proof that Euler's number e is transcendental.

1 Introduction

In 1966, Baker proved a landmark result about linear forms in the logarithms of algebraic numbers, which helped to earn him the Fields Medal in 1970. The following is a somewhat weak version of that result:

Baker's Theorem 1 (A. Baker, 1966). *Let $\alpha_1, ..., \alpha_n$ be non-zero algebraic numbers for which $\log \alpha_1, ..., \log \alpha_n, 2\pi i$ are linearly independent over $\mathbb{Q}$. Then*

$$\beta_1 \log \alpha_1 + ... + \beta_n \log \alpha_n \neq 0$$

for any algebraic numbers $\beta_1, ..., \beta_n$ that are not all zero.

This is weak in that, firstly, the theorem remains true if we just suppose that $\log \alpha_1, ..., \log \alpha_n$ are linearly independent over $\mathbb{Q}$, with no reference to $2\pi i$. (Baker asserted this in his original paper, and published a proof slightly later.) Moreover one can say, not just that the linear combination is non-zero, but that it is bounded away from zero in an effective way (as Baker did in his original paper). However, the theorem as stated admits a slightly more transparent proof.

Baker's theorem is ubiquitous in (transcendental) number theory, and has been discussed in many places, but it seems to be a commonly held view that the proof is rather mysterious and unintuitive. In this note we will attempt to dispel that view. To that end, we recall the following broad outline of a proof that Euler's number e is transcendental:

- Argue by contradiction, supposing that

  $$a_n e^n + a_{n-1} e^{n-1} + ... + a_1 e + a_0 = 0$$

  for some $n \in \mathbb{N}$ and $a_i \in \mathbb{Z}$ (with $a_n \neq 0$).

- Because of the properties of the exponential function, there is a large class of "nice" functions $F(i)$ that approximate e^i very well for $i = 0, 1, ..., n$. (For example, one can choose $F(i) = \sum_{k=0}^{m} f^{(k)}(i)$ for a suitable high degree polynomial f, where m is a parameter. Note that $F(i)$ satisfies $F'(i) \approx F(i)$ if f is chosen suitably.)

- For suitable choice of F, we can arrange that

  $$a_n F(n) + a_{n-1} F(n-1) + ... + a_1 F(1) + a_0 \approx a_n e^n + a_{n-1} e^{n-1} + ... + a_1 e + a_0 = 0$$

  is a *non-zero rational* with fairly small denominator. But it is clearly impossible to approximate zero very well by a non-zero rational with small denominator.

Note that the slightly complicated construction of an approximating function F replaces e.g. the appeal to the series expansion of e^x in the proof that e is irrational.

At a very high level, the proof that e is transcendental may be described in the following way: if e were algebraic, it would satisfy a "simple" polynomial equation (i.e. one of bounded degree and bounded height of coefficients), and this contradicts the analytic properties of the *function* e^x, because rational numbers are "fairly well spaced". We stress here that the key work in the proof is in determining functional properties of e^x.

We will see that, at this rather high level of inspection, the proof of Baker's Theorem is precisely analogous to the proof that e is transcendental. Firstly we suppose, for a contradiction, that

$$\beta_1 \log \alpha_1 + ... + \beta_n \log \alpha_n = 0$$

for some algebraic numbers $\beta_1, ..., \beta_n$ that are not all zero. In fact, without loss of generality (after possibly relabelling the α_i, and dividing through by a non-zero β_i) we may suppose that

$$\beta_1 \log \alpha_1 + ... + \beta_{n-1} \log \alpha_{n-1} - \log \alpha_n = 0.$$

We do not try to (directly) contradict this relation, which in particular is not a polynomial relation. Instead, going slightly beyond what happens in the proof that e is transcendental, we will indirectly construct a function $\phi(z)$ that vanishes, together with several of its derivatives, at many integer points z. The putative expression for $\log \alpha_n$ as a combination of $\log \alpha_1, ..., \log \alpha_{n-1}$ is input into this construction, and implies that $\phi(z)$ vanishes rather more than might be expected.

Having done this, we analyse the functional properties of $\phi(z)$. The so-called "extrapolation procedure" for doing this involves repeatedly playing off an analytic result, obtained by complex variable methods, against the fact (roughly— see §§$3-4$) that $\phi(z)$ takes algebraic values at integer z and therefore either vanishes at such z, or is effectively bounded away from zero there. The extrapolation method is, perhaps, the most novel ingredient of Baker's proof, and he wrote himself that it would "...probably be capable of considerable development for it applies in principle to many other auxiliary functions...".

Finally, the extrapolation procedure reveals that $\phi(z)$ must actually vanish at an enormous number of integer points. Because of the construction of $\phi(z)$, this implies a linear dependence over $\mathbb{Q}$ between $\log \alpha_1, ..., \log \alpha_n$ and $2\pi i$, which is a contradiction.

We conclude this introduction by recalling that the special case of Baker's theorem where $n = 2$ was proved rather earlier, by Gelfond and by Schneider independently in 1934. The author has not been able to view Schneider's argument, but it certainly seems fair to describe Baker's method as being a generalisation of Gelfond's method. In his book [3], Gelfond describes his argument as using "...the idea of analytic-arithmetic continuation." The author believes that to be a fitting description.

2 Construction of the auxiliary function $\phi(z)$

We now launch into the construction of the auxiliary function $\phi(z)$, which we want to vanish, along with many of its derivatives, at a number of integer points. Our account from this point on is a hybrid of Baker's original article [1], and Chapter 2 of his book [2] on this subject, with just a few changes to the exposition. We have a parameter $h \in \mathbb{R}$, which at the end of the proof will be taken to be large in a way depending on the α_i, the β_i, and n.

Recall the hypothesis that we wish to contradict, namely that

$$\beta_1 \log \alpha_1 + ... + \beta_{n-1} \log \alpha_{n-1} - \log \alpha_n = 0$$

for some algebraic numbers $\beta_1, ..., \beta_{n-1}$. Exponentiating, this becomes

$$\alpha_1^{\beta_1} ... \alpha_{n-1}^{\beta_{n-1}} \alpha_n^{-1} = 1,$$

but this is not much more helpful, because the purpose of proving the theorem is to understand how algebraic powers of algebraic numbers behave. On the other hand, we can say things about integer powers of algebraic numbers, which motivates the following choice of auxiliary function:

$$\phi(z) := \sum_{\lambda_1=0}^{L} ... \sum_{\lambda_n=0}^{L} p(\lambda_1, ..., \lambda_n) \alpha_1^{\lambda_1 z} ... \alpha_n^{\lambda_n z}, \quad z \in \mathbb{C},$$

where $L = [h^{2-1/(4n)}]$, and the $p(\lambda_1, ..., \lambda_n)$ are integers not all of which are zero, with absolute values at most e^{h^3}, such that

$$\frac{d^m}{dz^m} \phi(z) = 0 \quad \forall\, 0 \le m \le h^2,\ z \in \{1, 2, ..., h\}.$$

In a moment we will show that we *can* find such integers $p(\lambda_1, ..., \lambda_n)$, and the reader should note that the fact that we can do so with L smaller than h^2 by a power of h, which exploits our (to be contradicted) hypothesis about the α_i, is crucial to the subsequent argument.

Actually it is not too hard to find such integers, just a little fiddly. Because of our (to be contradicted) hypothesis, we can equivalently write

$$\phi(z) := \sum_{\lambda_1=0}^{L} ... \sum_{\lambda_n=0}^{L} p(\lambda_1, ..., \lambda_n) e^{(\lambda_1+\lambda_n\beta_1)z\log\alpha_1+...+(\lambda_{n-1}+\lambda_n\beta_{n-1})z\log\alpha_{n-1}},$$

and we will be done if we can choose the $p(\lambda_1, ..., \lambda_n)$ such that

$$\sum_{\lambda_1=0}^{L} ... \sum_{\lambda_n=0}^{L} p(\lambda_1, ..., \lambda_n)\alpha_1^{\lambda_1 z}...\alpha_n^{\lambda_n z}(\lambda_1+\lambda_n\beta_1)^{m_1}...(\lambda_{n-1}+\lambda_n\beta_{n-1})^{m_{n-1}}$$

vanishes for all $0 \leq m_1 + ... + m_{n-1} \leq h^2$ and all $z \in \{1, 2, ..., h\}$. Now if d is an upper bound for the degrees of the minimal polynomials of $\alpha_1, ..., \alpha_n, \beta_1, ..., \beta_{n-1}$, and a_1 is the leading coefficient in the minimal polynomial of α_1, we have for example that

$$(a_1\alpha_1)^j = \sum_{s=0}^{d-1} a_{1,s}^{(j)}\alpha_1^s \quad \forall\, j \in \mathbb{N} \cup \{0\},$$

for certain integers $a_{1,s}^{(j)}$. So multiplying through by $(a_1...a_n)^{Lz}b_1^{m_1}...b_{n-1}^{m_{n-1}}$, where a_i is the leading coefficient in the minimal polynomial of α_i, similarly for b_i, we see we would be done if

$$\sum_{s_1=0}^{d-1} ... \sum_{s_n=0}^{d-1}\sum_{t_1=0}^{d-1} ... \sum_{t_{n-1}=0}^{d-1} \alpha_1^{s_1}...\alpha_n^{s_n}\beta_1^{t_1}...\beta_{n-1}^{t_{n-1}} \Bigg(\sum_{\mu_1=0}^{m_1} ... \sum_{\mu_{n-1}=0}^{m_{n-1}} \sum_{\lambda_1=0}^{L} ... \sum_{\lambda_n=0}^{L} p(\lambda_1, ..., \lambda_n) \cdot$$

$$\cdot (\prod_{i=1}^{n} a_i^{Lz-\lambda_i z} a_{i,s_i}^{(\lambda_i z)})(\prod_{j=1}^{n-1} \binom{m_j}{\mu_j}(b_j\lambda_j)^{m_j-\mu_j}\lambda_n^{\mu_j} b_{j,t_j}^{(\mu_j)}) \Bigg)$$

vanished for all $0 \leq m_1 + ... + m_{n-1} \leq h^2$ and all $z \in \{1, 2, ..., h\}$.

But this can be achieved just by choosing the $p(\lambda_1, ..., \lambda_n)$ to solve a system of $M \leq (h^2+1)^{n-1}hd^{2n-1}$ integer linear equations, namely the equations that the inner bracket should equal zero for each choice of $m_1, ..., m_{n-1}, z, s_1, ..., s_n, t_1, ..., t_{n-1}$. If h is large enough, then we have $(L+1)^n \geq h^{2n-1/4} \geq 2M$ variables $p(\lambda_1, ..., \lambda_n)$, so we can certainly find a non-trivial solution. Moreover, an easy induction shows that $|a_{i,s}^{(j)}|, |b_{i,t}^{(j)}| \leq C^j$, where C depends on the α_i and β_i only; and therefore one has

$$|\prod_{i=1}^{n} a_i^{Lz-\lambda_i z} a_{i,s_i}^{(\lambda_i z)}| \leq K^{Lz} \leq K^{Lh}, \quad |\prod_{j=1}^{n-1} \binom{m_j}{\mu_j}(b_j\lambda_j)^{m_j-\mu_j}\lambda_n^{\mu_j} b_{j,t_j}^{(\mu_j)}| \leq (KL)^{h^2},$$

for a suitable constant K depending on the α_i, the β_i, and on n only. So the estimates on the size of the $p(\lambda_1, ..., \lambda_n)$ can be obtained using the following famous (but easily proved) result:

Siegel's Lemma 1 (C. Siegel, 1929, and others). *If $N > M > 0$ are integers, then the system of equations*

$$\sum_{j=1}^{N} u_{i,j} x_j = 0, \quad 1 \leq i \leq M$$

has a non-trivial solution in integers x_j with absolute values at most

$$1 + (N \max_{i,j} |u_{i,j}|)^{M/(N-M)}.$$

3 Easy estimates on $\phi(z)$

Having constructed $\phi(z)$, we begin to analyse its properties as a function of the complex variable z. In the first place, we have

$$\left| \log^{m_1} \alpha_1 ... \log^{m_{n-1}} \alpha_{n-1} \sum_{\lambda_1=0}^{L} ... \sum_{\lambda_n=0}^{L} p(\lambda_1, ..., \lambda_n) \alpha_1^{\lambda_1 z} ... \alpha_n^{\lambda_n z} (\lambda_1 + \lambda_n \beta_1)^{m_1} ... (\lambda_{n-1} + \lambda_n \beta_{n-1})^{m_{n-1}} \right|$$

$$\leq e^{h^3} (KL)^{h^2} K^{L|z|}, \quad 0 \leq m_1 + ... + m_{n-1} \leq h^2,$$

where K depends only on α_i, β_i, n, as before. It follows that, for $0 \leq m \leq h^2$ but for *all* $z \in \mathbb{C}$,

$$|\frac{d^m}{dz^m} \phi(z)| \leq K^{h^3 + L|z|}$$

for suitable (different) K depending on α_i, β_i, n only. Note the appearance of L, which we arranged to be a power of h slightly smaller than h^2, in this estimate.

Our other (fairly) easy estimate will encode, in a useful way, the fact that $\frac{d^m}{dz^m} \phi(z)$ takes (up to various multipliers involving the $\log \alpha_i$) algebraic values when z is an integer, and that algebraic numbers are fairly well spaced. Thus if $0 \leq m_1 + ... + m_{n-1} \leq h^2$, and $z \in \mathbb{N}$, then the number

$$X := (a_1 ... a_n)^{Lz} b_1^{m_1} ... b_{n-1}^{m_{n-1}} \sum_{\lambda_1=0}^{L} ... \sum_{\lambda_n=0}^{L} p(\lambda_1, ..., \lambda_n) \alpha_1^{\lambda_1 z} ... \alpha_n^{\lambda_n z} (\lambda_1 + \lambda_n \beta_1)^{m_1} ... (\lambda_{n-1} + \lambda_n \beta_{n-1})^{m_{n-1}}$$

is an algebraic integer with degree at most d^{2n-1} (by the Tower Law for field extensions, and since algebraic integers form a ring). Now arguing in a Liouville-esque way, either $X = 0$ or the norm of X is at least 1. But any conjugate of X has absolute value at most $K^{h^3 + Lz}$, arguing exactly as above, so either $X = 0$ or $|X| \geq K^{-d^{2n-2}(h^3 + Lz)}$. This obviously implies that for $0 \leq m \leq h^2$, and any $z \in \mathbb{N}$, we have

$$\frac{d^m}{dz^m} \phi(z) = \sum_{m_1 + ... + m_{n-1} = m} f_{m_1, ..., m_{n-1}}(z),$$

where we have

$$f_{m_1, ..., m_{n-1}}(z) = 0 \quad \text{or} \quad |f_{m_1, ..., m_{n-1}}(z)| \geq K^{-h^3 - Lz}$$

for suitable (different) K depending on α_i, β_i, n only.

4 The extrapolation argument

By construction, we know that $\phi(z)$ vanishes for $z \in \{1, 2, ..., h\}$. In this section we will argue, using the information that we also have about the derivatives of ϕ, and the bounds in §3, that actually $\phi(z)$ must vanish for $z \in \{1, 2, ..., (L+1)^n\}$ (or even for a larger set of z values). This will swiftly imply Baker's Theorem.

We need one key lemma, which is squarely complex-analytic. Rather than presenting a version that is highly tailored to our situation, it seems more revealing to state and prove a somewhat general version.

Baker's Lemma 1. *Let $f : \mathbb{C} \to \mathbb{C}$ be holomorphic, let $\epsilon > 0$, and let A, B, C, T, U be large real numbers. Suppose that $C \gg T/(A \log A) + UBA^{\epsilon}$, and that*

1. $|\frac{d^m}{dz^m} f(z)| \leq e^{T+U|z|} \quad \forall\, 0 \leq m \leq C,\ z \in \mathbb{C}$;
2. $\frac{d^m}{dz^m} f(z) = 0 \quad \forall\, 0 \leq m \leq C,\ z \in \{1, 2, ..., [A]\}$.

Then $|\frac{d^m}{dz^m} f(z)| \leq e^{-2(T+Uz)}$ for all $0 \leq m \leq C/2$ and all $z \in \{1, 2, ..., [AB]\}$.

We will prove this using the elegant argument from Baker's book [2], which exploits the maximum-modulus principle. However, we caution that this makes it appear that the exact vanishing of the derivatives in condition (2) is essential, and in fact that is not the case. (Indeed, the argument in Baker's paper [1], proving a quantitative version of his theorem, does not assume such vanishing.)

Fix any $0 \leq m \leq C/2$, and let $g(z) = \frac{d^m}{dz^m} f(z)$, so we are aiming to show that $|g(z)| \leq e^{-2(T+Uz)}$ for $z \in \{1, 2, ..., [AB]\}$. Because of assumption (2), we see that

$$\frac{g(z)}{(z-1)^{[C/2]}(z-2)^{[C/2]}...(z-[A])^{[C/2]}}$$

is a holomorphic function. Thus, by the maximum modulus principle applied on a circle about the origin with radius $A^{1+\epsilon}B$, for $z \in \{1, 2, ..., [AB]\}$ we have

$$|g(z)| \leq \max_{|w|=A^{1+\epsilon}B} \left(|g(w)| \frac{|z-1|^{[C/2]}...|z-[A]|^{[C/2]}}{|w-1|^{[C/2]}...|w-[A]|^{[C/2]}} \right) \leq e^{-(\epsilon/2)\log A[A][C/2]} \max_{|w|=A^{1+\epsilon}B} |g(w)|.$$

Using the assumption (1), and that $AC \gg (T/\log A) + UA^{1+\epsilon}B$, the result follows.

Q.E.D.

This lemma is applied to the functions $f_{m_1,...,m_{n-1}}$ appearing at the end of §3, where the hypotheses are satisfied with $T = h^3 \log K$, $U = L \log K$, $B = h^{1/8n}$, etc. (The reader should note that the various estimates we derived for ϕ in previous sections were actually derived separately for each function $f_{m_1,...,m_{n-1}}$). Because of the dichotomy established at the end of §3, the conclusion of the lemma is untenable unless the derivatives of $f_{m_1,...,m_{n-1}}$ vanish for z on the wide range supplied. Thus, iteratively applying the lemma $O(n^2)$ times, the claim that

$$\phi(z) = 0 \quad \forall\, z \in \{1, 2, ..., (L+1)^n\}$$

follows.

5 Conclusion of the proof

Recall that we had

$$\phi(z) = \sum_{\lambda_1=0}^{L} ... \sum_{\lambda_n=0}^{L} p(\lambda_1, ..., \lambda_n)\alpha_1^{\lambda_1 z} ... \alpha_n^{\lambda_n z}, \quad z \in \mathbb{C},$$

and we now know that $\phi(1) = \phi(2) = ... = \phi((L+1)^n) = 0$. This means that the vector of numbers $p(\lambda_1, ..., \lambda_n)$ is a non-trivial element of the kernel of a certain $(L+1)^n \times (L+1)^n$ matrix, which must therefore have determinant zero.

But this matrix is clearly a Vandermonde matrix, so the vanishing of its determinant implies that

$$\alpha_1^{\lambda_1} ... \alpha_n^{\lambda_n} = \alpha_1^{\lambda'_1} ... \alpha_n^{\lambda'_n}$$

for some distinct tuples of integers $(\lambda_1, ..., \lambda_n)$ and $(\lambda'_1, ..., \lambda'_n)$. This contradicts the assumption that $\log \alpha_1, ..., \log \alpha_n, 2\pi i$ are linearly independent over $\mathbb{Q}$.

Q.E.D.

Again, it is perhaps worth pointing out that there are other ways to end this proof that are more "robust" than relying on the appearance of a Vandermonde determinant. In contrast, the extrapolation argument set out in §4 is fundamental to Baker's approach.

References

[1] A. Baker. Linear forms in the logarithms of algebraic numbers. *Mathematika* **13**, pp 204-216. 1966

[2] A. Baker. *Transcendental number theory.* Reissue published by Cambridge University Press. 1990

[3] A. O. Gelfond. *Transcendental and Algebraic Numbers.* Translated by L. Boron. Dover publ., New York. 1960