A Brief but Historic Article of Siegel

Rodrigo A. Pérez

À Adrien Douady et les hiboux

The two papers published by Carl L. Siegel in 1942 were printed on ten consecutive pages in the October issue of the *Annals of Mathematics*. In one of these papers Siegel gave the first positive solution to a small denominator problem, and by doing so he showed that there was hope for a successful attack on one of the most important problems of the previous sixty years. It was a remarkable achievement that has earned [15] acclaim as

> "one of the landmark papers of the twentieth century."[1]

To justify this high opinion, we need to understand Siegel's proof and its historical background. In this article I will explain

- What small denominators are and why they are important.
- The linearization problem and its status in 1942.
- Siegel's original proof, including the correction of a minor gap.
- Some of the major mathematical developments in the wake of [15].

How to Read This Article

Siegel was a master of concise writing. In only six pages he included a motivation for the result and presented an intricate yet self-contained proof. One drawback of his exposition is that it takes considerable ingenuity to see how all the pieces fit together, even though the only prerequisite is to be able to compute the radius of convergence of a power series using the root test.

The initial aim of this article was simply to give an easy-to-read account of the original proof, but soon I found myself tracking the ideas that must have gone into play as Siegel found his arguments and prepared them for publication. As a result, the proof I first produced is now substantially simplified and annotated. Motivated undergraduates with a semester of analysis under their belts should be able to follow the entire argument. Graduate students writing a paper for the first time may find it interesting to pursue a comparative reading of this material and [15], which is widely available through JSTOR. As a note of warning for them, I kept Siegel's notation for the most part but made some changes (particularly regarding subindices) that simplify the exposition and keep compatibility between sections.

The following two sections give some historical background on small denominators. The next section explains the problem, and the subsequent section describes diophantine conditions. The theorem and its proof span the remaining sections. Note that the two lemmas are numbered as in [15]; nevertheless, Lemma 1 is proved at the end because it is only incidental to the main argument.

The remarks at the end give a minimal account of events after the publication of [15]. The reader interested only in the historical aspects of this story can safely skip the section on diophantine conditions and the last five sections.

What Is a Small Denominator?

Before we can answer this question, consider a harmonic oscillator $\ddot{x} + \omega_1^2 x = 0$, whose general

Rodrigo A. Pérez is assistant professor in the Department of Mathematical Sciences at Indiana University–Purdue University Indianapolis. His email address is rperez@math.iupui.edu.

Work supported by NSF grant DMS-0701557.

[1] *The quote is from* [3, p. 482]. *The present title is derived from similar praise in* [18, p. 6].

http://www-history.mcs.st-and.ac.uk/

C. L. Siegel

solution $x(t) = A\cos(\omega_1 t + \varphi)$ represents a periodic motion of frequency ω_1. If we add a periodic perturbation of frequency $\omega_2 \neq \omega_1$,

$$\ddot{x} + \omega_1^2 x + \cos(\omega_2 t) = 0,$$

the new solution has the form

$$x(t) = A\cos(\omega_1 t + \varphi) - \frac{\cos(\omega_2 t)}{\omega_1^2 - \omega_2^2}.$$

This function is only periodic when ω_2 is a rational multiple of ω_1, but even when that is not the case, it features nice, bounded, quasi-regular oscillations. Note that if ω_2 is close to $\pm\omega_1$, the quotient on the right can become arbitrarily large. To see what happens as ω_2 approaches ω_1, let us focus on the simplest initial conditions $x(0) = \dot{x}(0) = 0$. Then

$$x(t) = \frac{\cos(\omega_1 t) - \cos(\omega_2 t)}{\omega_1^2 - \omega_2^2},$$

and L'Hôpital's rule gives $x(t) = \frac{-t\sin(\omega_1 t)}{2\omega_1}$ in the limit as $\omega_2 \to \omega_1$. This last function is unbounded because the periodic kicks of the perturbation build up without canceling. This is the essence of the phenomenon of *resonance* which is so troublesome to engineers.

A striking example of amplitude growth near resonance is given by the tide system in the Bay of Fundy, Nova Scotia: tides are caused by the gravitational influence of the moon and the sun on big bodies of water. The largest component of this influence is the *principal lunar semidiurnal constituent* whose frequency is 12.42 hours (one half of the average time needed for the Earth to rotate once relative to the moon). This is very close to the 13.3 ± 0.4 hours needed by large waves to travel from the mouth of the bay to the inner shore and back [7]. Combined with a host of secondary effects, this match of frequencies produces the highest tides in the world.

A more sophisticated phenomenon than simple resonance occurs when two distinct periodic motions of frequencies ω_1 and ω_2 interact with each other. If nonlinear terms are present, the perturbation is often expressed by a power series whose coefficients have terms $m\omega_1 + n\omega_2$ ($m, n \in \mathbb{Z}$) in the denominators. We say that ω_1 and ω_2 display a *near resonance* whenever the linear combination $m\omega_1 + n\omega_2$ is unusually small. Now we can describe the small denominator problem, which is simply(!) to establish convergence of the series on the face of multiple near resonances that may yield big coefficients.

Small denominators are found most commonly in the perturbative theory of hamiltonian mechanics. The prototypical setting is the mutual perturbation of two planets around the sun. Indeed, the question of stability in the solar system, which eluded Poincaré, was the initial motivation for Siegel's work.

Celestial Mechanics

In celestial mechanics, small denominators are linked to long-term irregularities in planetary orbits. Here, "long term" is meant relative to the period of the orbits. The earliest observed instance of this phenomenon is known as the *great inequality of Jupiter and Saturn.*

The mean motions of Jupiter and Saturn (the average angle they cover daily in their orbits around the sun) are $\omega_1 = 299.1283''$ per day and $\omega_2 = 120.4548''$ per day, respectively. Since ω_1/ω_2 is so close to $5/2$, the time needed by Jupiter to complete five orbits around the sun (21662.945 days) is nearly identical to the time taken by Saturn to complete two orbits (21518.440 days).

After a conjunction occurs (i.e., Jupiter aligns between Saturn and the sun), it takes 21760.362 days for Jupiter to cover $8°5'40''$ in excess of five orbits and for Saturn to cover the same $8°5'40''$ in excess of two orbits, thus reaching alignment again. Two other conjunctions occur at one third and two thirds of this time interval. As a consequence, the perturbation exerted by the two planets on each other, which is largest near conjunction, tends to build up around three equally spaced regions that slowly advance along the orbits. The cumulative effect goes though a cycle of about 918 years, during which Jupiter is displaced as far as $48.5'$ and Saturn as far as $21'$ from their undisturbed trajectories.

This discrepancy was noticed in actual observations and troubled astronomers during the better parts of the seventeenth and eighteenth centuries. It was finally explained by Laplace in a famous three-part memoir published between 1784 and 1786. The first exposition of the methods of

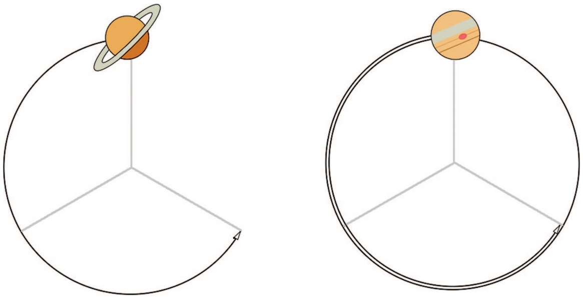

Figure 1. If the ratio of periods was exactly 2/5, Saturn would cover two thirds of its orbit in the time that Jupiter covers five thirds of its own.

perturbation theory is also due to Laplace and appeared in the first two volumes of his *Mécanique céleste*, published in 1799. By the second half of the nineteenth century, perturbation theory had been developed to a very high degree. In 1860–67, C.-E. Delaunay published two 900-page volumes [6] in which he computed the orbit of the moon under the perturbative influence of the sun. The three resulting series for the moon's latitude, longitude, and parallax include all terms up to order 7 and span 121 pages altogether.

Chapter III of Delaunay's opus contains the first analytic description of small denominators [6, pg. 87]. The hamiltonian of the perturbed motion of the moon is a function R of the mean motions ω_M, ω_S of the moon and the sun (n and n' in [6]), and two other astronomical quantities. When the periodic component of R is written explicitly, the solutions to Hamilton's equations feature trigonometric series with linear combinations of ω_M and ω_S in the denominators. Delaunay pointed out that because of this, higher order terms can be larger than first-order terms, making a truncated approximation useless.

Although small denominators show up in other contexts in celestial mechanics, the underlying setting is always a series of the form

$$\sum_{m\in(\mathbb{Z}^n)^*} a_m \frac{e^{i(m\cdot\omega)t}}{m\cdot\omega},$$

where ω is a vector of frequencies. If ω has many near resonances, the coefficients may grow too large too often, threatening the convergence of the series. H. Poincaré was the first to recognize and address this difficulty. In an often quoted fragment of the *Méthodes Nouvelles* [14, §148–149], he admitted that his methods did not guarantee the convergence of these series, but he granted the (remote) possibility that some particular convergent cases may exist. Poincaré was remarkably prescient in guessing both the existence of solutions and the difficulty of the proofs.

The Linearization Problem

Small denominators appear in many other settings in which irrational frequencies resonate. Siegel focused on a model problem in which no physical considerations obscure the small-denominator issue.

Let $f(z) = \sum_{r=1}^{\infty} a_r z^r$ be a nonlinear complex analytic function with a fixed point at 0. The value $f'(0)$ is called the *multiplier* of 0 and will be denoted λ. Here, λ is assumed different from 0.

The linearization problem asks if there is a function $\varphi(z) = \sum_{k=1}^{\infty} c_k z^k$ satisfying

$$\varphi(\lambda z) = (f\circ\varphi)(z). \tag{1}$$

Note that if such a map exists, multiplication by a constant c before applying φ simply rescales the domain, so $z \mapsto \varphi(cz)$ is also a linearizing map. By setting $c = 1/\varphi'(0)$, the coefficient c_1 can be assumed to be 1.

The Kœnigs-Poincaré theorem [12, p. 77] guarantees a solution to the linearization equation (1) whenever $|\lambda| \neq 1$. If $\lambda^n = 1$, an easy computation shows that f is linearizable if and only if $f^n = \mathrm{id}$. This leaves $\lambda = e^{2\pi i\theta}$ with irrational θ as the most interesting case, and this condition will be assumed from now on. Geometrically, (1) says that f is conjugate to an irrational rotation around the fixed point. The maximal domain of linearization is known today as a *Siegel disk*.

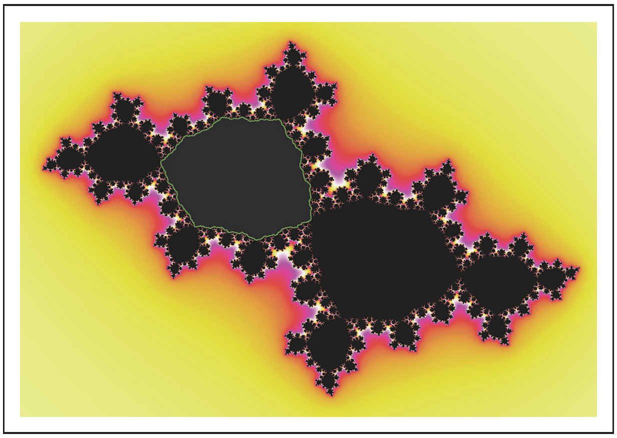

Figure 2. The Julia set of $f(z) = z^2 + c$, where c was chosen so that f has a Siegel disk with rotation number $\theta = (\sqrt[3]{2}-1) \approx 0.259921\ldots$ The disk is the large highlighted region in the center.

In terms of the power series of f and φ, equation (1) has the form

$$\sum_{k=1}^{\infty} c_k(\lambda z)^k = \sum_{r=1}^{\infty} a_r \left(\sum_{\ell=1}^{\infty} c_\ell z^\ell\right)^r;$$

or, singling out the first term on the right (note that $a_1 = \lambda$),

$$\text{(2)} \qquad \sum_{k=2}^{\infty} c_k(\lambda^k - \lambda)z^k = \sum_{r=2}^{\infty} a_r \left(\sum_{\ell=1}^{\infty} c_\ell z^\ell \right)^r .$$

Since c_1 is taken to be 1, equation (2) gives a concrete recursive description of the sequence $\{c_k\}$. It states that $c_k(\lambda^k - \lambda)z^k$ is the sum of all z^k-monomials present in the right-hand side. Now, the expression $a_r\left(\sum_{\ell=1}^{\infty} c_\ell z^\ell\right)^r$ produces z^k-monomials exactly when $2 \le r \le k$. These monomials have the form $a_r \cdot (c_{\ell_1} z^{\ell_1}) \cdot \ldots \cdot (c_{\ell_r} z^{\ell_r})$, where the powers of z add up to k, so for $k \ge 2$,

$$\text{(3)} \qquad c_k = \left(\frac{1}{\lambda^k - \lambda} \right) \left(\sum_{r=2}^{k} \sum_{\ell_1 + \ldots + \ell_r = k} a_r \cdot c_{\ell_1} \cdot \ldots \cdot c_{\ell_r} \right).$$

Equation (3) seems to solve the linearization problem as it defines explicitly the coefficients of a power series solution to (1). However, depending on the value of λ, the absolute values

$$\text{(4)} \qquad \varepsilon_k := \frac{1}{|\lambda^{k+1} - \lambda|} = \frac{1}{|\lambda^k - 1|}$$

(note the index discrepancy)

can get very large very often. This threatens the convergence of the power series for φ, and, indeed, it is possible that the series solution is only formal. We will call ε_k an SD-*term*, leaving the reader at liberty to decide whether SD stands for "Small Denominator" or "Siegel Disk".

Poincaré's opinion on the scarcity of linearizable maps was prevalent for half a century. In 1917 G. Pfeiffer [13] constructed the first nonlinearizable example, and in 1928 H. Cremer [5] found a dense G_δ set of angles for which no rational function is linearizable. In this atmosphere, Siegel's paper came as a surprise, as he found a large family of angles θ (satisfying condition (5)) for which linearization is possible. To state his result we need the following definition and a discussion of diophantine conditions.

Definition. The power series of f is convergent, so there is a smallest $a > 0$ satisfying $|a_r| \le a^{r-1}$ for all r. If f is replaced by the conjugate function $af(z/a)$, the multiplier remains intact, but we can assume $|a_r| \le 1$. Such f is said to be *normalized*.

Diophantine Conditions

Siegel required that λ satisfy

$$\text{(5)} \qquad \log|\lambda^n - 1| = \mathcal{O}(\log n) \text{ as } n \to \infty.$$

This says that there is a constant $\hat{\nu} > 0$ such that for sufficiently large n,

$$\big|\log|\lambda^n - 1|\big| \le \hat{\nu} \log n.$$

Since $\log|\lambda^n - 1|$ is at most $\log 2$ (for λ^m accumulating at -1), condition (5) makes sense as a bound on $-\log|\lambda^n - 1|$ when $\lambda^n - 1$ becomes *small*. Taking exponentials gives $|\lambda^n - 1|^{-1} \le n^{\hat{\nu}}$ for n larger than some M. This can be changed to a condition for *every* n by letting K be the larger of 1 and $\max_{n \le M}\left\{(n^{\hat{\nu}}|\lambda^n - 1|)^{-1}\right\}$. Then, setting $\nu = \hat{\nu} + \log_2 K$ gives[2]

$$\text{(6)} \qquad |\lambda^n - 1|^{-1} \le K n^{\hat{\nu}} < (2n)^{\nu}.$$

Recall that $\lambda = e^{2\pi i\theta}$, so $|\lambda^n - 1| = 2|\sin(\pi n\theta)|$. More precisely [12, p. 129], if m is the nearest integer to $n\theta$, i.e., if $|n\theta - m| < 1/2$, then $|\lambda^n - 1| = 2\sin(\pi|n\theta - m|)$, and since the graph of $\sin(\pi x)$ lies between the lines $y = 2x$ and $y = \pi x$ when $0 \le x \le 1/2$,

$$4|n\theta - m| \le |\lambda^n - 1| < 2\pi|n\theta - m|.$$

Thus, (6) is equivalent to

$$\text{(7)} \qquad \left|\theta - \frac{m}{n}\right| > \frac{Q}{n^{\nu+1}},$$

which is the defining property of a *diophantine number* θ *of order* $\nu + 1$. Denote the set of all θ satisfying (7) by $\mathcal{D}(\nu + 1)$. It turns out that the lowest ν such that $\mathcal{D}(\nu + 1)$ is nonempty is $\nu = 1$, and that $\mathcal{D}(2)$ has measure 0. On the other hand, $\bigcap_{\nu > 2} \mathcal{D}(\nu)$ has full measure on $[0, 1]$. This implies that λ satisfies (5) with probability one.

The Theorem

With the above notation, Siegel's result can be stated in its full strength as follows:

Theorem. *Given $\nu \ge 1$, let $b_\nu = (3 - \sqrt{8})/2^{5\nu+1}$. If $\theta \in \mathcal{D}(\nu + 1)$, then all normalized f of multiplier $\lambda = e^{2\pi i\theta}$ are linearizable, and the linearizing map φ has radius of convergence at least b_ν.*

In other words, all normalized f with multiplier λ in a subset of full measure of the circle have a guaranteed radius of linearization. The bound depends only on the diophantineness order of λ and is given explicitly.

The Majorant Method

To prove that the radius of convergence of φ is positive, Siegel used Cauchy's majorant method to estimate the exponential rate of growth of the sequence $\{c_k\}$. This requires several steps. First, note from (3) that the absolute values $|c_k|$ are bounded by the real sequence $\{\widehat{c_k}\}$ defined by $\widehat{c_1} = 1$ and

$$\text{(8)} \qquad \widehat{c_k} = \varepsilon_{k-1} \left(\sum_{r=2}^{k} \sum_{\ell_1 + \ldots + \ell_r = k} \widehat{c_{\ell_1}} \cdot \ldots \cdot \widehat{c_{\ell_r}} \right)$$

(recall the normalization assumption $|a_r| \le 1$).

[2] *See the comment following the proof of Lemma 2 for a justification of this choice of logarithmic base.*

The structure of these numbers is somehow obscured by their recursive definition. Following (8), the first coefficients are

$$\begin{aligned}
\hat{c}_2 &= \varepsilon_1([\hat{c}_1\hat{c}_1]) = \varepsilon_1\,,\\
\hat{c}_3 &= \varepsilon_2([\hat{c}_1\hat{c}_2 + \hat{c}_2\hat{c}_1] + [\hat{c}_1\hat{c}_1\hat{c}_1]) = 2\varepsilon_2\varepsilon_1 + \varepsilon_2\,,\\
\hat{c}_4 &= \varepsilon_3([\hat{c}_1\hat{c}_3 + \hat{c}_2\hat{c}_2 + \hat{c}_3\hat{c}_1] + [\hat{c}_1\hat{c}_1\hat{c}_2 + \hat{c}_1\hat{c}_2\hat{c}_1\\
&\qquad\qquad + \hat{c}_2\hat{c}_1\hat{c}_1] + [\hat{c}_1\hat{c}_1\hat{c}_1\hat{c}_1])\\
&= 4\varepsilon_3\varepsilon_2\varepsilon_1 + 2\varepsilon_3\varepsilon_2 + \varepsilon_3\varepsilon_1\varepsilon_1 + 3\varepsilon_3\varepsilon_1 + \varepsilon_3.
\end{aligned}\tag{9}$$

Thus, $\hat{c}_k$ is the sum of many expressions, each of which is the product of several SD-terms (not necessarily different). It is possible to describe explicitly which products appear, but that is not important here and will be omitted.

Let τ_k be the number of products in the expansion of $\hat{c}_k$, and δ_k the maximum of their values. From (9), it is clear for instance that $\tau_4 = 4 + 2 + 1 + 3 + 1 = 11$, but the precise collection of SD-terms whose product realizes the maximum δ_k will depend on λ. Note, however, that (setting aside the factor ε_{k-1} of $\hat{c}_k$) the largest product of SD-terms in the expansion of $\hat{c}_k$ appears in some product $\widehat{c_{\ell_1}} \cdot \ldots \cdot \widehat{c_{\ell_r}}$ and therefore has to be the product of the largest products in each of $\widehat{c_{\ell_1}}, \ldots, \widehat{c_{\ell_r}}$. In other words, δ_k is given by $\delta_1 = 1$ and

$$\delta_k = \varepsilon_{k-1} \cdot \max_{\substack{\ell_1+\ldots+\ell_r=k\\ 2\le r\le k}} \{\delta_{\ell_1} \cdot \ldots \cdot \delta_{\ell_r}\} \qquad (k \ge 2). \tag{10}$$

This recursive definition allows for a much more efficient computation of δ_k.

Cauchy's majorant method is based on the obvious fact that

$$|c_k| \le \hat{c}_k \le \delta_k \tau_k, \tag{11}$$

and on the observation that the values τ_k are given by $\tau_1 = 1$ and the recursion

$$\tau_k = \left(\sum_{r=2}^{k} \sum_{\ell_1+\ldots+\ell_r=k} \tau_{\ell_1} \cdot \ldots \cdot \tau_{\ell_r}\right)$$

(compare with (8)).

The theorem will follow from (11) and the exponential bounds (12) on $\{\tau_k\}$ and (23) on $\{\delta_k\}$. It is useful to keep in mind that the definition of the numbers τ_k is related to the structure of the linearization equation (1), while the definition of δ_k reflects the effect of the angle θ on the SD-terms ε_k.

τ_k Grows Exponentially

The *Schröder numbers* $\{\tau_k\}$ are well known in combinatorics (see [16, sequence A001003], [17], and references therein). The initial values are 1, 1, 3, 11, 45, 197, 903, 4279, 20793, 103049, ..., and their generating function $y(x) = \sum \tau_\ell x^\ell$ satisfies the functional equation

$$y = x + \sum_{r=2}^{\infty} y^r$$

essentially by the same line of reasoning that produces (3) from (2). Since the above is just $y = x + \frac{y^2}{1-y}$, it follows that

$$y(x) = \frac{1 + x - \sqrt{1 - 6x + x^2}}{4},$$

so the radius of convergence of y is the absolute value of the smallest root of $1 - 6x + x^2$; i.e., $(3 - \sqrt{8})$. In particular, the sequence $\{\tau_k\}$ grows as a power of $(3 - \sqrt{8})^{-1} = (3 + \sqrt{8})$. More accurately, it is known [11] that it has the asymptotic behavior

$$\tau_k \sim \frac{W(3 + \sqrt{8})^k}{k^{3/2}}, \tag{12}$$

where $W = \frac{1}{4}\sqrt{(\sqrt{18} - 4)/\pi} = 0.069478\ldots$

The Subtle Estimate

The bulk of Siegel's proof is concerned with showing that the sequence $\{\delta_k\}$ defined by recursion (10) has an exponential bound whenever the SD-terms satisfy (6). Using the notation in (4), the diophantine condition reads

$$\varepsilon_k \le (2k)^\nu. \tag{13}$$

This will be called the *basic estimate*. Since each δ_k is a product of $\mathcal{O}(k)$ SD-terms, (13) is far from giving an efficient bound on the growth of the sequence. Siegel's insight, and one of the reasons his result was so influential, was the realization that once an SD-term is large, it takes several steps before another SD-term can have comparable size. This is made precise in the following argument. Since

$$\lambda^q(\lambda^{p-q} - 1) = (\lambda^p - 1) - (\lambda^q - 1),$$

and $|\lambda^q| = 1$, we get via the triangle inequality,

$$|\lambda^{p-q} - 1| \le |\lambda^p - 1| + |\lambda^q - 1|.$$

In SD-notation the above reads

$$\varepsilon_{p-q}^{-1} \le \varepsilon_p^{-1} + \varepsilon_q^{-1} \le 2(\min\{\varepsilon_p, \varepsilon_q\})^{-1}.$$

Then, applying the basic estimate (13) to ε_{p-q},

$$\min\{\varepsilon_p, \varepsilon_q\} \le 2^{\nu+1}(p - q)^\nu. \tag{14}$$

This is much better than the trivial $\min\{\varepsilon_p, \varepsilon_q\} \le \min\{(2p)^\nu, (2q)^\nu\}$ and will be called the *subtle estimate*. Siegel must have been pleased with the simplicity of this core idea, because he actually allowed himself a small boasting note at this point [15, p. 610]:

> "This simple remark is the main argument of the whole proof."

A Bound on a Product of SD-Terms

How does an estimate on the *least* of two SD-terms yield an upper estimate on a product of SD-terms? The following proof of Lemma 2 reformulates the inductive argument in [15] to answer this question. To simplify notation, let $N = 2^{2\nu+1}$ (a constant that depends on the diophantineness value).

Lemma 2. *Given $r+1$ indices $k_0 > \ldots > k_r \geq 1$, the following holds:*

$$\prod_{p=0}^{r} \varepsilon_{k_p} < N^{r+1} \cdot k_0^{\nu} \prod_{p=1}^{r} (k_{p-1} - k_p)^{\nu}. \tag{15}$$

Note that it is the indices, rather than the SD-terms themselves, that are arranged by size in (strictly) descending order.

Proof. The proof is by induction. The basic estimate (13) covers the case $r = 0$. If $r = 1$, the basic and subtle estimates give

$$\begin{aligned} \varepsilon_{k_0} \cdot \varepsilon_{k_1} &\leq (2^{\nu} \max\{k_0^{\nu}, k_1^{\nu}\}) \cdot (2^{\nu+1}|k_0 - k_1|^{\nu}) \\ &< N^2 \cdot k_0^{\nu} \cdot |k_0 - k_1|^{\nu}. \end{aligned}$$

Now consider the case of $r+1 \geq 3$ SD-terms, and let ε_{k_j} be the smallest one. By the induction hypothesis, the remaining SD-terms satisfy (15) with the index k_j missing. If $j = 0$, then $\varepsilon_{k_j} = \varepsilon_{k_0}$ is bounded by

$$2^{\nu+1}(k_0 - k_1)^{\nu} < N(k_0 - k_1)^{\nu}$$

and (15) holds. A similar argument applies when $j = r$.

When $0 < j < r$, the inductive bound on $\varepsilon_{k_0} \cdot \ldots \cdot \varepsilon_{k_{j-1}} \cdot \varepsilon_{k_{j+1}} \cdot \ldots \cdot \varepsilon_{k_r}$ contains the factor $(k_{j-1} - k_{j+1})^{\nu}$. Let a, b be such that $\{k_a, k_b\} = \{k_{j-1}, k_{j+1}\}$ and $|k_a - k_j| \geq |k_j - k_b|$. It follows that $(k_{j-1} - k_{j+1}) \leq 2|k_a - k_j|$, while (14) gives $\varepsilon_{k_j} \leq 2^{\nu+1}|k_j - k_b|^{\nu}$. Then the product $\prod_{p=0}^{r} \varepsilon_{k_p}$ of all $r+1$ SD-terms is bounded by

$$\begin{aligned} &N^r \cdot k_0 \cdot (k_0 - k_1)^{\nu} \cdot \ldots \cdot (k_{j-1} - k_{j+1})^{\nu} \cdot \varepsilon_{k_j} \cdot \ldots \cdot (k_{r-1} - k_r)^{\nu} \\ &\leq N^r \cdot k_0 \cdot (k_0 - k_1)^{\nu} \cdot \ldots \cdot 2^{\nu}|k_a - k_j|^{\nu} \cdot 2^{\nu+1}|k_j - k_b|^{\nu} \cdot \ldots \cdot (k_{r-1} - k_r)^{\nu} \\ &= N^{r+1} \cdot k_0 \cdot (k_0 - k_1)^{\nu} \cdot \ldots \cdot (k_{j-1} - k_j)^{\nu} \cdot (k_j - k_{j+1})^{\nu} \cdot \ldots \cdot (k_{r-1} - k_r)^{\nu}. \square \end{aligned}$$

It is worth noting that in the factorization $N = 2^{\nu} \cdot 2^{\nu+1}$ used in the last equality, both powers of 2 come from different sources. The factor 2^{ν} is due to the fact that the interval $[k_{j-1}, k_{j+1}]$ is shorter than *twice* the longer of $[k_{j-1}, k_j]$ and $[k_j, k_{j+1}]$. The factor $2^{\nu+1}$ on the other hand comes from (6), where logarithmic base 2 was chosen simply so that N has a clean expression.

δ_k Grows Exponentially

The stage is set to find an exponential bound on δ_k. It may be impossible to reconstruct what Siegel did to discover a proof, but here is a plausible scenario. Start by writing $\delta_k \leq AC^k/k^B$ with the intention of exploiting the recursive decomposition (10) to find values A, B, C that make the inequality true. The case $\delta_1 \leq AC$ suggests setting $A = C^{-1}$ and solving

$$\delta_k \leq \frac{C^{k-1}}{k^B} \tag{16}$$

for all k. A solution to (16) can always be upgraded to one that satisfies

$$B \overset{(a)}{>} 0 \text{ and } C \overset{(b)}{\geq} 2^B$$

by a suitable increase in C. The extra conditions (a) and (b) have the advantage that

$$\begin{aligned} \delta_{j_1} \cdot \delta_{j_2} &\leq \frac{C^{j_1+j_2-2}}{j_1^B \cdot j_2^B} = C^{-1}\left(\frac{1}{j_1} + \frac{1}{j_2}\right)^B \frac{C^{j_1+j_2-1}}{(j_1+j_2)^B} \\ &\overset{(a)}{\leq} C^{-1} 2^B \frac{C^{j_1+j_2-1}}{(j_1+j_2)^B} \overset{(b)}{\leq} \frac{C^{j_1+j_2-1}}{(j_1+j_2)^B}, \end{aligned}$$

and more generally,

$$\delta_{j_1} \cdot \ldots \cdot \delta_{j_t} \leq \frac{C^{J-1}}{J^B}, \tag{17}$$

whenever $j_1 + \ldots + j_t = J$.

Now suppose we are in possession of numbers B and C that satisfy (a), (b), and (16) for all $k \geq 1$ smaller than k_0. Then inequality (19) below can be reached by the following argument:

In the decomposition (10) of k_0 the sum of indices of all deltas is equal to k_0. In particular, there can be at most one index larger than $k_0/2$. If this is the case, write $\delta_{k_0} = \varepsilon_{k_0-1} \cdot \delta_{k_1} \cdot \Delta_1$ where $k_1 > k_0/2$, and consider the decomposition (10) of k_1. There may still be an index larger than k_0. If so, write $\delta_{k_1} = \varepsilon_{k_1-1} \cdot \delta_{k_2} \cdot \Delta_2$, and continue this process until the decomposition of some δ_{k_r} has no delta with index larger than k_0. This produces the following tower

$$\begin{aligned} \delta_{k_0} &= \varepsilon_{k_0-1} \cdot \delta_{k_1} \cdot \Delta_1\,, \\ \delta_{k_1} &= \varepsilon_{k_1-1} \cdot \delta_{k_2} \cdot \Delta_2\,, \\ &\vdots \\ \delta_{k_{r-1}} &= \varepsilon_{k_{r-1}-1} \cdot \delta_{k_r} \cdot \Delta_r\,, \\ \delta_{k_r} &= \varepsilon_{k_r-1} \cdot \delta_{\ell_1} \cdot \ldots \cdot \delta_{\ell_s}\,, \end{aligned} \tag{18}$$

where $k_0 > k_1 > \ldots > k_r > k_0/2$. Note that each Δ_p lumps together many unnamed deltas. Their indices add up to $k_{p-1} - k_p$, and are all at most k_0. The indices ℓ_q add up to k_r and are also at most k_0. These conditions on indices will be necessary in order to apply Lemma 1 in (22).

Let us collapse the tower by repeated substitution to get

$$\delta_{k_0} = \prod_{p=0}^{r} \varepsilon_{k_p-1} \cdot \prod_{p=1}^{r} \Delta_p \cdot (\delta_{\ell_1} \cdot \ldots \cdot \delta_{\ell_s}).$$

Then, applying Lemma 2 to the product of SD-terms, inequality (17) to each Δ_p, and inequality

(16) to each δ_{ℓ_q} gives

(19)
$$\begin{aligned}\delta_{k_0} &\leq \left[N^{r+1}\cdot k_0^{\nu}\prod_{p=1}^{r}(k_{p-1}-k_p)^{\nu}\right]\\ &\quad\cdot\left[\prod_{p=1}^{r}\frac{C^{(k_{p-1}-k_p)-1}}{(k_{p-1}-k_p)^B}\right]\cdot\left[\prod_{q=1}^{s}\frac{C^{\ell_q-1}}{\ell_q^B}\right]\\ &= (N^{r+1}\cdot C^{-r-s})\\ &\quad\cdot\left(k_0^{\nu}\cdot\prod_{p=1}^{r}(k_{p-1}-k_p)^{\nu-B}\cdot\prod_{q=1}^{s}\ell_q^{-B}\right)\cdot C^{k_0}.\end{aligned}$$

Recall that this inequality is contingent on finding B, C that satisfy (a), (b), and (16) with $1 \leq k < k_0$. The goal is to discover a smart choice of B and C so that the right-hand side of (19) is also bounded by $\frac{C^{k_0-1}}{k_0^B}$. Accordingly, the next step is to find a way to extract a factor k_0^{-B} from the middle parenthesis. More precisely, we look for an auxiliary inequality of the form

$$\left(\prod_{p=1}^{r}(k_{p-1}-k_p)^{\nu-B}\cdot\prod_{q=1}^{s}\ell_q^{-B}\right)\leq k_0^{-\nu-B}\cdot\Xi, \tag{20}$$

where Ξ may depend on r and s, but not on k_0. Since the sum of factors $\sum(k_{p-1}-k_p)+\sum\ell_q$ is k_0, a simple heuristic for an inequality like (20) to be possible is that the sums of exponents on the left and right sides balance out. In the present case this means $(\nu - B) - B = -\nu - B$, or

$$B \overset{(c)}{=} 2\nu.$$

This is compatible with condition (a). As it turns out, the heuristic works, and Siegel found the inequality (24) of Lemma 1 (see the following section). Together with (c) and the index conditions mentioned after (18), inequality (24) yields the following version of (20):

(21)
$$\left(\prod_{p=1}^{r}(k_{p-1}-k_p)^{-\nu}\cdot\prod_{q=1}^{s}\ell_q^{-2\nu}\right)\leq\left(\frac{k_0}{2^{r+s-1}}\right)^{-3\nu},$$

which means that (19) is bounded by

$$\begin{aligned}&(N^{r+1}\cdot C^{-r-s})\cdot\left(k_0^{\nu}\cdot\left(\frac{k_0}{2^{r+s-1}}\right)^{-3\nu}\right)\cdot C^{k_0}\\ &\quad=\left((2^{2\nu+1})^{r+1}\cdot(2^{3\nu})^{r+s-1}\cdot C^{-r-s+1}\right)\cdot\frac{C^{k_0-1}}{k_0^{2\nu}}\\ &\qquad\leq\left(\frac{2^{2\nu+1}\cdot 2^{3\nu}}{C}\right)^{r+s-1}\cdot\frac{C^{k_0-1}}{k_0^{2\nu}}.\end{aligned}\tag{22}$$

The last expression is smaller than $\frac{C^{k_0-1}}{k_0^{2\nu}}$ when $C \geq 2^{5\nu+1}$. This last condition is compatible with (b) and (c); and so, substituting $B = 2\nu$ and $C = 2^{5\nu+1}$ in (19) and (22) yields a proof that

$$\delta_{k_0}\leq\frac{(2^{5\nu+1})^{k_0-1}}{k_0^{2\nu}}\tag{23}$$

for all $k_0 \geq 1$, and the theorem is proved.

The Auxiliary Inequality

Although the proof is correct, the middle inequality in line (8) of [15] does not hold when k is odd and $t = \left\lceil\frac{k}{2}\right\rceil$. The following proof of Lemma 1 simplifies Siegel's exposition and avoids this minor lapse by considering instead inequality (29), which calls for a separate treatment of the cases $t = 2$ and $k = 2, 3, 4$.

Observation 1. The cubic polynomial $P(x) = (R - x)(x - S)^2$ with $R > S$ has derivative $P'(x) = (2R + S - 3x)(x - S)$, so S is a critical point. The second derivative is $P''(x) = 2R + 4S - 6x$, which, evaluated at $x = S$, is $2R - 2S > 0$. Thus S is the only local minimum of P. It follows that if the interval $I = [a, b]$ lies to the right of S, then

$$\min_{x\in I}\{P(x)\} = \min\{P(a), P(b)\}.$$

Lemma 1. *Let three integers $k \geq 2$, $r \geq 0$, and $s \geq 2$ be given. If the integers $x_1, \ldots, x_r$ and $y_1, \ldots, y_s$ belong to $\left\{1, \ldots, \left\lfloor\frac{k}{2}\right\rfloor\right\}$ and satisfy $\sum_{p=1}^{r} x_p + \sum_{q=1}^{s} y_q = k$ with $\sum_{q=1}^{s} y_q > k/2$, then*

$$\prod_{p=1}^{r}x_p\cdot\prod_{q=1}^{s}y_q^2\geq\left(\frac{k}{2^{r+s-1}}\right)^3.\tag{24}$$

Proof. Let $t = r + s \geq 2$. Since $2t - 2 \leq 2^{t-1}$, it suffices to prove

$$\prod x_p\cdot\prod y_q^2\geq\left(\frac{k}{2t-2}\right)^3.\tag{25}$$

Some cases are immediate. If $t = 2$, then $r = 0$ and $s = 2$, so $y_1 = y_2 = k/2$ and (25) holds. Also, (25) holds trivially when $k \leq 2t - 2$; this is the case for $k = 2, 3, 4$ when $t \geq 3$. It remains to consider what happens when $t \geq 3$, $k \geq 5$, and $k > 2t - 2$, or equivalently,

$$3\leq t\leq\left\lceil\frac{k}{2}\right\rceil.\tag{26}$$

The smallest product $\prod x_p$ is realized when $r - 1$ factors are equal to one, and the remaining factor is what is left of $\sum x_p$. Thus, $\prod x_p$ has the lower bound $\sum x_p - (r - 1)$. Analogously, the product $\prod y_q$ can be estimated from below by $\sum y_q - (s - 1)$. However, a sharper bound is available when $\sum y_q - (s - 1) > \left\lfloor\frac{k}{2}\right\rfloor$, for in that case the least product is realized by $s - 2$ factors equal to one, a factor equal to $\left\lfloor\frac{k}{2}\right\rfloor$, and a factor equal to the rest (so that no y_q is larger than $\left\lfloor\frac{k}{2}\right\rfloor$). In short,

(27)
$$\begin{aligned}&\prod x_p\geq\sum x_p - r + 1,\\ &\prod y_q\geq\begin{cases}\sum y_q - s + 1 & \text{if } \sum y_q - s + 1\leq\left\lfloor\frac{k}{2}\right\rfloor,\\ \left(\sum y_q - s - \left\lfloor\frac{k}{2}\right\rfloor + 2\right)\cdot\left\lfloor\frac{k}{2}\right\rfloor & \text{if } \sum y_q - s + 1\geq\left\lfloor\frac{k}{2}\right\rfloor.\end{cases}\end{aligned}$$

The sum $\sum y_q$ can take values between $\left(\left\lfloor \frac{k}{2}\right\rfloor + 1\right)$ and $(k - r)$. The analysis that follows breaks into two cases.

Case 1: If $\left(\left\lfloor \frac{k}{2}\right\rfloor + 1\right) \le \sum y_q \le \left(\left\lfloor \frac{k}{2}\right\rfloor + s - 1\right)$, then (27) gives

(28)
$$\prod x_p \cdot \prod y_q^2 \ge (k - \sum y_q - r + 1) \cdot (\sum y_q - s + 1)^2 .$$

Let $R = k - r + 1$ and $S = s - 1$, so (28) reads $\prod x_p \cdot \prod y_q^2 \ge P\left(\sum y_q\right)$. Now, $R > S$ so Observation 1 applies. The product $\prod x_p \cdot \prod y_q^2$ is bounded from below by the minimum of P in the range of $\sum y_q$. Since the range is included in the (larger) interval $I = \left[\left(\left\lfloor \frac{k}{2}\right\rfloor - r + 1\right), \left(\left\lfloor \frac{k}{2}\right\rfloor + s - 1\right)\right]$, and $\left(\left\lfloor \frac{k}{2}\right\rfloor - r + 1\right) > S$ by (26), the product $\prod x_p \cdot \prod y_q^2$ is bounded by the least of

$$P\left(\left\lfloor \frac{k}{2}\right\rfloor - r + 1\right) = \left\lceil \frac{k}{2}\right\rceil \left(\left\lfloor \frac{k}{2}\right\rfloor - t + 2\right)^2$$

and

$$P\left(\left\lfloor \frac{k}{2}\right\rfloor + s - 1\right) = \left(\left\lceil \frac{k}{2}\right\rceil - t + 2\right) \left\lfloor \frac{k}{2}\right\rfloor^2 .$$

But

$$\begin{aligned}&\left\lceil \frac{k}{2}\right\rceil \cdot \left(\left\lfloor \frac{k}{2}\right\rfloor - t + 2\right)^2 \le \\ &\quad \left\lceil \frac{k}{2}\right\rceil \cdot \left(\left\lfloor \frac{k}{2}\right\rfloor - t + 2\right) \cdot \left(\left\lceil \frac{k}{2}\right\rceil - t + 2\right) \le \\ &\qquad \left(\left\lfloor \frac{k}{2}\right\rfloor + 1\right) \cdot \left(\left\lfloor \frac{k}{2}\right\rfloor - 1\right) \cdot \left(\left\lceil \frac{k}{2}\right\rceil - t + 2\right)\end{aligned}$$

because $t \ge 3$. The last line is smaller than $\left(\left\lceil \frac{k}{2}\right\rceil - t + 2\right) \cdot \left\lfloor \frac{k}{2}\right\rfloor^2$, so for $\sum y_q$ in this range,

$$\prod x_p \cdot \prod y_q^2 \ge \left\lceil \frac{k}{2}\right\rceil \left(\left\lfloor \frac{k}{2}\right\rfloor - t + 2\right)^2 .$$

Case 2: If $\left(\left\lfloor \frac{k}{2}\right\rfloor + s - 1\right) \le \sum y_q \le (k - r)$, then (27) gives

$$\prod x_p \cdot \prod y_q^2 \ge (k - \sum y_q - r + 1) \cdot \left(\sum y_q - s - \left\lfloor \frac{k}{2}\right\rfloor + 2\right)^2 \cdot \left\lfloor \frac{k}{2}\right\rfloor^2 .$$

In this case, let $R = k - r + 1$ and $S = s + \left\lfloor \frac{k}{2}\right\rfloor - 2$, so (26) implies $R > S$. Now the range of $\sum y_q$ is in the interval $I = \left[\left(\left\lfloor \frac{k}{2}\right\rfloor + s - 1\right), (k - r)\right]$, which clearly lies to the right of S. Observation 1 bounds $\prod x_p \cdot \prod y_q^2$ from below by the least of

$$P\left(\left\lfloor \frac{k}{2}\right\rfloor + s - 1\right) = \left(\left\lceil \frac{k}{2}\right\rceil - t + 2\right) \cdot \left\lfloor \frac{k}{2}\right\rfloor^2$$

and

$$P(k - r) = \left(\left\lceil \frac{k}{2}\right\rceil - t + 2\right)^2 \cdot \left\lfloor \frac{k}{2}\right\rfloor^2 .$$

Obviously the former is smaller, so $\prod x_p \cdot \prod y_q^2 \ge \left(\left\lceil \frac{k}{2}\right\rceil - t + 2\right) \cdot \left\lfloor \frac{k}{2}\right\rfloor^2$; but at the end of Case 1 this was shown to be larger than $\left\lceil \frac{k}{2}\right\rceil \left(\left\lfloor \frac{k}{2}\right\rfloor - t + 2\right)^2$, so

(29)
$$\prod x_p \cdot \prod y_q^2 \ge \left\lceil \frac{k}{2}\right\rceil \cdot \left(\left\lfloor \frac{k}{2}\right\rfloor - t + 2\right)^2 \ge \frac{k}{2} \cdot \left(\frac{k-1}{2} - t + 2\right)^2$$

for all valid values of $\sum y_q$.

Now, $(k-1)/2 - t + 2$ is a linear function of t, while $(k-1)/(2t-2)$ is convex. Since the former is larger than the latter when $t = 3$ and when $t = \left\lceil \frac{k}{2}\right\rceil$, the same inequality is valid in the full range of t, so continuing from (29),

$$\begin{aligned}&\frac{k}{2} \cdot \left(\frac{k-1}{2} - t + 2\right)^2 \ge \frac{k}{2} \cdot \left(\frac{k-1}{2t-2}\right)^2 = \\ &\quad (t-1) \cdot \left(\frac{k-1}{k}\right)^2 \cdot \left(\frac{k}{2t-2}\right)^3 \ge \\ &\quad 2 \cdot \left(\frac{4}{5}\right)^2 \cdot \left(\frac{k}{2t-2}\right)^3 > \left(\frac{k}{2t-2}\right)^3 . \quad \square\end{aligned}$$

Conclusion

The influence of [15] was due in part to the elementary nature of the majorant method, which I hope to have conveyed; but its major impact was conceptual. Although the small-denominator problem he solved was simpler than those found in celestial mechanics, Siegel's proof showed that the convergence issue could be handled. His main observation was that it is possible to quantify how frequently small denominators of comparable size can appear. This was a fruitful idea that he extended to similar problems in several variables.

Moreover, the notion that number theory was relevant to dynamical systems (via diophantine approximations) served as a catalyst for much ensuing research. In time, the study of rotation domains became a subject of its own, with major contributions by A. Brjuno, T. Cherry, M. Herman, J.-C. Yoccoz, R. Pérez-Marco, M. Shishikura, C. McMullen, X. Buff, and A. Chéritat. In this area, much of the effort was directed to geometric considerations; for instance, the behavior of critical points near the boundary of a Siegel disk or the existence of Siegel disks with smooth boundary.

Alas, the approach in [15] did not apply in the setting of hamiltonian dynamics where techniques were most urgently sought. In 1954, at the International Congress of Mathematicians in Amsterdam, A. Kolmogorov announced a theorem (inspired in part by Siegel) that would change the face of dynamical systems in the form of KAM theory:

In an *integrable* hamiltonian system the phase space is foliated by invariant tori, each with an associated *rotation vector* $\omega \in \mathbb{R}^d$. The solutions within a given torus are conjugate to the linear translation $p \mapsto p + \omega t$. When the entries of ω are rationally independent, the solutions are dense in the torus and are said to be *quasi-periodic*.

Prior to 1954 it was expected that a small perturbation of an integrable system would destroy this structure, so that most trajectories would break away from their original tori and wander around phase space. That is, the assumption was that the perturbed system should be ergodic.

Kolmogorov showed that the structure of the perturbed solutions is much more interesting than that. While many solutions do wiggle about phase space, diophantine[3] rotation vectors still give rise to (deformed) invariant tori where solutions are quasi-periodic. Thus the regions of chaotic and regular behavior are inextricably blended together, and each has positive measure. In other words, the ergodic hypothesis has to be discarded.

Notice how this result takes us from studying the convergence of individual series to a global study of the space of solutions. The theory does not verify convergence for individual initial conditions, but rather guarantees a positive probability of convergence, while making clear the role of the diophantine condition. In a sense, the small-denominator problem has been bypassed.

After Kolmogorov's announcement, techniques like the majorant method were abandoned (even by Siegel) in favor of global analysis in the KAM spirit. One of the few people to revisit Siegel's method was A. Brjuno [1]. He improved the original proof and described (what Yoccoz [18] would later prove is) the largest class of angles θ for which every analytic function f with fixed point 0 of multiplier $e^{2\pi i\theta}$ is linearizable. References to a few other applications of the majorant method can be found in [4].

Many questions remain open. Are there bounded Siegel disks whose boundary is not a Jordan curve? What is the structure around Cremer points, where linearization is impossible? In 2005 X. Buff and A. Chéritat completed a project, started by A. Douady in the 1990s, to construct polynomial Julia sets of positive measure. The strategy is to approximate a Cremer polynomial (whose Julia set has no interior) by a sequence of linearizable polynomials while delicately controlling the reduction in area of the corresponding Siegel disks. Their results [2] promise new life for an interesting subject.

Acknowledgments

I am indebted to M. Aspenberg for the long conversations that led to this paper, to R. Roeder for many illuminating suggestions, to J. Smillie for editorial advice, and to the many people that have read or listened to portions of this material.

Note: Figure 2 was created with FractalStream, a research-oriented program used to explore dynamical systems. Available at `http://code.google.com/p/fractalstream`.

References

[1] Alexander D. Brjuno, Analytic form of differential equations, I, II, *Trudy Moskov. Mat. Obšč.* **25** (1971), 119–262; ibid. **26** (1972), 199–239.

[2] Xavier Buff and Arnaud Chéritat, Ensembles de Julia quadratiques de mesure de Lebesgue strictement positive, *C. R. Math. Acad. Sci. Paris* **341**(11) (2005), 669–674.

[3] Xavier Buff, Christian Henriksen, and John H. Hubbard, Farey curves, *Experiment. Math.* **10**(4) (2001), 481–486.

[4] Luigi Chierchia and Corrado Falcolini, Compensations in small divisor problems, *Comm. Math. Phys.* **175**(1) (1996), 135–160.

[5] Hubert Cremer, Zum Zentrumproblem, *Math. Ann.* **98**(1) (1928), 151–163.

[6] Charles-E. Delaunay, 1867, Théorie du Mouvement de la Lune, 2 Vols. in *Mem. Acad. Sci.* 28 and 29 (Mallet-Bachelier, Paris, 1860, and Gauthier-Villars, Paris, 1867).

[7] Christopher Garrett, Tidal resonance in the Bay of Fundy and Gulf of Maine, *Nature* **238**(5365) (1972), 441–443.

[8] Boris Hasselblatt and Anatole Katok, The development of dynamics in the 20th century and the contribution of Jürgen Moser, *Ergodic Theory Dynam. Systems* **22**(5) (2002), 1343–1364.

[9] Michael-R. Herman, Recent results and some open questions on Siegel's linearization theorem of germs of complex analytic diffeomorphisms of $\mathbf{C}^n$ near a fixed point, *VIIIth International Congress on Mathematical Physics (Marseille, 1986)*, World Sci. Publishing, Singapore, 1987, pp. 138–184.

[10] John H. Hubbard, The KAM theorem, *Kolmogorov's Heritage in Mathematics*, Springer, Berlin, 2007, pp. 215–238.

[11] Donald E. Knuth, *The Art of Computer Programming*, Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 2nd edition, 1975, Volume 1: Fundamental algorithms, Addison-Wesley Series in Computer Science and Information Processing.

[12] John Milnor, *Dynamics in One Complex Variable*, volume 160 of Annals of Mathematics Studies, Princeton University Press, Princeton, NJ, 3rd edition, 2006.

[13] G. A. Pfeiffer, On the conformal mapping of curvilinear angles. The functional equation $\varphi[f(x)] = a_1\varphi(x)$, *Trans. Amer. Math. Soc.* **18**(2) (1917), 185–198.

[14] Henri Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste. Tome II. Méthodes de MM. Newcomb, Gyldén, Lindstedt et Bohlin*, Dover Publications, New York, NY, 1957.

[15] Carl L. Siegel, Iteration of analytic functions, *Ann. of Math.* **43**(2) (1942), 607–612.

[16] Neil J. A. Sloane, (2006), *The On-Line Encyclopedia of Integer Sequences*, `http://www.research.att.com/~njas/sequences/`.

[17] Richard P. Stanley, Hipparchus, Plutarch, Schröder, and Hough, *Amer. Math. Monthly* **104**(4) (1997), 344–350.

[18] Jean-C. Yoccoz, *Petits diviseurs en dimension 1*, Société Mathématique de France, Paris, 1995, Astérisque No. 231 (1995).

[3] *Compare the condition* $|\omega\cdot\nu| \geq \frac{\kappa}{|\nu|^\tau}$ *(for all* $\nu \in \mathbb{Z}^d\setminus\{0\}$*) with* (7).