Expo. Math. 4 (1986), 97–109

On invariant manifolds of complex analytic mappings near fixed points

JÜRGEN PÖSCHEL

1 Statement of the Result

We consider a complex analytic diffeomorphism g in $\mathbb{C}^n$ in the neighbourhood of a fixed point p, which we may place at the origin. We assume that the linearization of g at p is diagonizable. Then, possibly after a linear change of coordinates, we can write

$$g: \quad z \mapsto \Lambda z + \hat{g}(z),$$

where

$$\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n),$$

and $\hat{g}$ vanishes up to first order at $0 \in \mathbb{C}^n$.

The linear mapping $z \mapsto \Lambda z$ has a very simple structure. For instance, for any subset $\lambda_1, \ldots, \lambda_s$ of eigenvalues with $1 \leq s \leq n$, the corresponding eigenspace obviously is an invariant manifold M, on which the mapping is linear with these eigenvalues.

We consider the following question. Under which conditions does such an invariant manifold persist after restoring the nonlinearity $\hat{g}$? More precisely, under which conditions does there exist locally a complex analytic, invariant manifold $\mathfrak{M}$ of g, which is tangent to M, and on which the restricted map is analytically equivalent to its linear part?

This generalizes the question, under which conditions the map g can be linearized at the fixed point.

A unique *formal* power series representation of $\mathfrak{M}$ is easily found by comparison of coefficients, provided we have

$$\lambda_1^{k_1} \cdot \dots \cdot \lambda_s^{k_s} - \lambda_i \neq 0, \qquad k_1 + \dots + k_s \geq 2, \quad 1 \leq i \leq n,$$

with nonnegative integers $k_1, \dots, k_s$. In usual multiindex notation this reads

$$\lambda^k - \lambda_i \neq 0, \qquad |k| \geq 2, \quad 1 \leq i \leq n, \tag{1}$$

where

$$\lambda = (\lambda_1, \dots, \lambda_s),$$

and $k = (k_1, \dots, k_s)$, $|k| = k_1 + \dots + k_s$. See section 5 for the details.

The question of *convergence* is far more subtle. The differences (1) enter into the denominators of the coefficients of the formal solution. Unless they are uniformly bounded away from zero, they give rise to *small divisors*. It is well known [8, 22] that they may cause the divergence of the formal solution, if they approach zero too rapidly. To obtain convergence, we therefore have to put suitable bounds on them.

Define the function ω by

$$\omega(m) = \min_{2 \leq |k| \leq m} \min_{1 \leq i \leq n} \left| \lambda^k - \lambda_i \right|, \qquad m \geq 2. \tag{2}$$

We call the divisors $\lambda^k - \lambda_i$ *admissible*, if

$$\sum_{\nu \geq 0} q_\nu^{-1} \log \omega^{-1}(q_{\nu+1}) < \infty \tag{3}$$

for some sequence of integers $1 = q_0 < q_1 < \dots$. This in fact holds if and only if

$$\sum_{\nu \geq 0} 2^{-\nu} \log \omega^{-1}(2^{\nu+1}) < \infty.$$

For the simple proof see Brjuno [6]. For instance, divisors are admissible which satisfy

$$\left| \lambda^k - \lambda_i \right| \geq c \, |k|^{-N}, \qquad |k| \geq 2, \quad 1 \leq i \leq n,$$

for some $c > 0$ and some large N. But much worse bounds are also admissible – see for example Rüssmann [17].

Our result can now be stated as follows.

Theorem *If the divisors* (1) *are admissible, then there exists locally a complex analytic invariant manifold* $\mathfrak{M}$ *of* g*, which is tangent to the eigenspace* M *of* $\lambda_1, \ldots, \lambda_s$*, and on which the mapping is analytically equivalent to its linear part.*

We will prove the theorem by the traditional majorant method following Siegel [20, 21] and Brjuno [6]. In the case of full linearization ($s = n$) the crucial observation is that only one of two divisors

$$\left|\lambda^k - \lambda_i\right|, \quad \left|\lambda^l - \lambda_j\right|$$

can really be small if $|k - l|$ is small. This, however, is not necessarily true in the case of partial linearization ($s < n$) if $i \neq j$. But we observe that actually we only need to compare divisors for which $i = j$. Then the same arguments apply.

Nowadays small divisor problems are usually approached by the more versatile iteration method of Newton type introduced by Kolmogorov, Arnold and Moser [1, 13, 14]. As we will indicate, however, the result in the above form does not seem to be within reach of this method. Stronger assumptions seem to be necessary.

2 Historical Notes

In the case $n = 1$, hence $s = n$, we are confronted with the problem of linearizing an analytic map in the complex plane in the neighbourhood of a fixed point. This leads to the functional equation of Schröder [19]

$$g \circ \psi = \psi \circ \lambda,$$

where λ denotes multiplication with the single eigenvalue λ of Λ. It has a unique formal solution $\psi = z + \ldots$, the "Schröder series", if λ is not a root of unity.

The convergence of this formal solution was already known to Poincaré [16] and proven by Koenigs [12] for $0 < |\lambda| \neq 1$. In this case, the fixed point is unstable, and no small divisors occur. On the other hand, Cremer [8] showed that on the unit circle $|\lambda| = 1$, there exists a dense set of λ, for which the Schröder series diverges for suitable choices of g. These λ are not roots of unity, but some $\left|\lambda^k - 1\right|$ get small very fast as k tends to infinity. See [22] for a simple example of that kind.

In 1942, Siegel [20] was the first to overcome the difficulty of the small divisors using the majorant method in an ingenious way. He proved convergence of the Schröder series provided λ satisfies

$$\left|\lambda^k - 1\right| \geq c k^{-N}, \qquad k \geq 1,$$

for some $c > 0$ and some large N. This condition is satisfied on a set of full measure on the unit circle. Later on, he proved an analogous linearization result for complex analytic vector fields of arbitrary dimensions [21]. This result was generalized to mappings of arbitrary dimensions by Sternberg [23]. Finally, Brjuno [5, 6] slightly varied the majorant method to allow for a wider class of small divisor estimates. He introduced the condition (3), which seems to be the weakest small divisor condition necessary to prove convergence. Incidentally, already in 1964, Cherry [7] constructed a one dimensional example indicating the necessity of condition (3). Unfortunately, his argument seems to be incorrect.

The majorant method is confined to complex analytic problems, and does not apply to real analytic problems in general, not to mention the just differentiable ones. These became tractable only with the introduction of a rapidly converging iteration method of Newton type by Kolmogorov, Arnold and Moser [1, 13, 14]. Of course, this technique also applies to complex analytic problems. Rüssmann [17] treated the Schröder equation this way, incorporating the more general small divisor estimates of Brjuno. Arnold [2] indicated how to deal with higher dimensional mappings along the same lines. A detailed exposition can be found in [3].

Soon after, the new approach was cast into the form of various generalized implicit function theorems on function spaces. See Zehnder [25] and Hamilton [10] for recent examples. In this manner, Sternberg [24] again considered the problem of linearizing a mapping, but he had to assume

$$|\lambda_i| \leq 1, \qquad 1 \leq i \leq n,$$

for some purely technical reason. He himself expressed doubt that this condition was necessary. Gray [9] then removed it, following essentially Sternberg's exposition. Finally, Zehnder [26] gave a concise proof in terms of an implicit function theorem.

As to full linearization the result can actually be sharpened. Condition (1) is necessary to ensure that *every* map g can be linearized formally. However, it may happen that condition (1) is violated by some choices of k and i (possibly infinitely often), and still there exists a formal transformation which linearizes g. Rüssmann [18] showed that in this case there also exists a convergent transformation, provided all the *nonvanishing* divisors (1) are admissible. See also Rüssmann [18] for an important application of that fact.

The problem of partial linearization ($s < n$) seems to have attained only little attention. Brjuno [6] studies rather extensively the existence of invariant manifolds at equilibria of analytic vector fields in the presence of small divisors. See also Bibikov [4]. The case of a mapping was considered in the Diplomarbeit of Klingenberg [11]. They all use the iteration method.

This note was actually prompted by the paper of Klingenberg. It turned out, that in the case of partial linearization, the majorant method does give a stronger result, as we will indicate now.

3 Comparison with the KAM method

The KAM method applies a succession of coordinate changes, which in the limit is supposed to transform the given map into a "normal form"

$$h\colon \quad z \mapsto \Lambda z + \hat{h}(z),$$

such that, for $z = (u, v) \in \mathbb{C}^s \times \mathbb{C}^{n-s}$,

$$\hat{h}\Big|_{v=0} = 0.$$

In these coordinates, $\mathfrak{M} = \{v = 0\}$.

To assure convergence of this scheme, however, it is not enough to ask that $\hat{h}|_{v=0} = 0$. In addition, the derivatives of $\hat{h}$ in the normal direction of $\mathfrak{M}$ have to be considered leading to a more specific "normal form". Accordingly, the transformations employed have to contain a term, which is linear in the normal coordinate v and allows to keep control over these normal directions.

This phenomenon is rather well known for the problem of constructing invariant manifolds at stationary points of vector fields corresponding to purely imaginary eigenvalues. See for instance Brjuno [6], who introduces the notion of a "complete normal form of an invariant surface", or Bibikov [4], who refers to "quasi-normal forms". See also Moser [15] for the closely related problem of perturbing a not normally hyperbolic invariant torus carrying quasi-periodic motions. Klingenberg [11] approaches the problem for mappings in the same manner.

The upshot is that normal derivatives introduce the additional divisors

$$\lambda^k \lambda_i - \lambda_j, \qquad |k| \geq 1, \quad s + 1 \leq i, j \leq n. \tag{4}$$

Of these, only finitely many can vanish because of (1). The others have to be admissible to render a convergent iteration scheme.

We now give an example, where the divisors (1) are very well admissible, while the divisors (4) are not. In other words, the admissibility of only the divisors (1) is not sufficient to apply the KAM method.

The following example is taken from Klingenberg [11].

Take $s = 1$, $n = 3$, and write

$$\lambda_j = e^{2\pi i\theta_j}, \qquad 1 \le j \le 3,$$

with $0 \le \theta_j \le 1$. We also drop the index 1 to simplify the notation. We have, for integer k and $j = 1, 2, 3$,

$$\left|\lambda^k - \lambda_j\right| \sim \left\|k\theta - \theta_j\right\|,$$
$$\left|\lambda^k\lambda_i - \lambda_j\right| \sim \left\|k\theta - (\theta_i - \theta_j)\right\|,$$

where $\|\xi\| = \min_{k\in\mathbb{Z}} |\xi - k|$, and the tilde indicates that either side is bounded from above and below by a constant multiple of the other side.

We choose θ so that

$$\omega(m) = \min_{1\le k\le m} \|k\theta\|$$

defines an admissible function ω. That is, ω satisfies (3). Then we pick μ in $[0, 1]$ such that

$$\left|[k\theta] - \mu\right| \le \frac{1}{k!}$$

for infinitely many k, where $[\xi]$ denotes the fractional part of a real number ξ. For instance, we can take

$$\mu \in \bigcap_{\nu\ge 0} I_\nu,$$

where the intervals I_ν are constructed inductively as follows. Set $I_0 = [0, 1]$. If closed intervals $I_0 \supset I_1 \supset \cdots \supset I_\nu$ of positive length have already been constructed, choose k so that $[k\theta] \in I_\nu$, and set

$$I_{\nu+1} = \left\{\xi : \left|\xi - [k\theta]\right| \le 1/k!\right\} \cap I_\nu.$$

This interval has positive length. The intersection of this nested sequence of closed intervals is not empty.

Now the measure of the two sets of ξ in $[0, 1]$ which satisfy

$$\min_{1\le k\le m} \|k\theta - \xi\| \ge \omega(m), \qquad m \ge 2,$$

and

$$\min_{1\le k\le m} \|k\theta - (\xi - \mu)\| \ge \omega(m), \qquad m \ge 2,$$

respectively, is greater than $1/2$, if, say, $\omega(m) \le 10^{-1}m^{-2}$. We can therefore pick ξ in the intersection of these two sets and set

$$\theta_2 = \xi, \qquad \theta_3 = \xi - \mu.$$

With this choice of $\theta, \theta_2, \theta_3$, we have

$$\min_{1\le k\le m} \left\|k\theta - \theta_j\right\| \ge \omega(m), \qquad m \ge 2, \quad 1 \le j \le 3.$$

But the differences

$$\|k\theta - (\theta_2 - \theta_3)\| = \|k\theta - \mu\| \le \left|[k\theta] - \mu\right|$$

are certainly not admissible, because they roughly decay like $1/k!$.

4 An example

We give a simple example of a map which can be linearized on some invariant manifold, but not in a full neighbourhood of a fixed point.

Consider the case where all eigenvalues of Λ are powers of a fixed complex number:

$$\lambda_i = \mu^{q_i}, \qquad 1 \le i \le n,$$

where

$$q_1 = \cdots = q_s = q > q_{s+1} \ge \cdots \ge q_n$$

are integers with $q \ge 1$ and $1 \le s < n$, and μ is a complex number on the unit circle such that

$$\omega(m) = \min_{1\le l\le m} \left|\mu^l - 1\right| \tag{5}$$

is admissible.

Then

$$\left|\lambda^k - \lambda_i\right| = \left|\mu^{|k|q} - \mu^{q_i}\right| = \left|\mu^{|k|q-q_i} - 1\right|.$$

For $|k| \geq 2$ we have $|k|q - q_i \geq |k|q - q \geq 1$, so we can apply (5) to obtain

$$\left|\lambda^k - \lambda_i\right| \geq \omega(|k|q - q_i) \geq \omega(|k|Q), \qquad |k| \geq 2,$$

for some constant Q. It follows that the small divisors are admissible, and the Theorem applies. On the other hand, a full linearization is in general not possible, because (1) is violated in that case.

5 Proof of the Theorem

We propose to obtain the manifold $\mathfrak{M}$ by an analytic embedding

$$\psi: \quad w \mapsto z = Jw + \hat{\psi}(w) \tag{6}$$

of a ball at the origin in $\mathbb{C}^s$ into $\mathbb{C}^n$. The $s \times n$-matrix J maps $\mathbb{C}^s$ into the eigenspace M of $\lambda_1, \ldots, \lambda_s$, and $\hat{\psi}$ vanishes up to first order at the origin.

The embedding ψ has to satisfy the equation

$$g \circ \psi = \psi \circ \Lambda_s,$$

where $\Lambda_s = \operatorname{diag}(\lambda_1, \ldots, \lambda_s)$. The linear terms already agree, so we have to find a solution $\hat{\psi}$ of the nonlinear equation

$$\hat{\psi} \circ \Lambda_s - \Lambda\hat{\psi} = \hat{g} \circ \psi, \tag{7}$$

which is complex analytic in a neighbourhood of the origin in $\mathbb{C}^s$.

The formal solution is straightforward. Write

$$\hat{\psi} = \sum_{|k|\geq 2} \psi_k w^k, \qquad \psi_k \in \mathbb{C}^n,$$

where $k = (k_1, \ldots, k_s)$, and

$$\hat{g} = \sum_{|l|\geq 2} g_l z^l, \qquad g_l \in \mathbb{C}^n,$$

where $l = (l_1, \ldots, l_n)$. Then (7) becomes

$$\sum_{|k|\geq 2} E_k \psi_k w^k = \sum_{|l|\geq 2} g_l \Bigl(\sum_{|m|\geq 1} \psi_m w^m \Bigr)^l, \tag{8}$$

where

$$E_k = \lambda^k I - \Lambda,$$

and the first order coefficients ψ_k, $|k| = 1$, are determined by (6). This equation allows us to determine the coefficients ψ_k for $|k| \geq 2$ uniquely by recursion, since by assumption,

$$\det E_k = \prod_{i=1}^{n} \bigl(\lambda^k - \lambda_i\bigr) \neq 0.$$

To prove convergence of this formal solution in a neighbourhood of the origin, we have to show that

$$\sup_k \frac{1}{|k|} \log |\psi_k| < \infty. \tag{9}$$

We can assume, possibly after stretching the z-variables, that g is analytic and bounded on $|z| = \max_i |z_i| < 1$. Then

$$|g_l| \leq M, \qquad |l| \geq 2.$$

It follows from (8) that then

$$|\psi_k| \leq \varepsilon_k^{-1} M \sum_{\substack{k_1+\cdots+k_\nu=k \\ \nu\geq 2}} \bigl|\psi_{k_1}\bigr| \cdot \ldots \cdot \bigl|\psi_{k_\nu}\bigr|, \qquad |k| \geq 2,$$

where

$$\varepsilon_k = \min_{1\leq i\leq n} \bigl|\lambda^k - \lambda_i\bigr| = \bigl\| E_k^{-1} \bigr\|^{-1},$$

and the sum is taken over all possible decompositions of k into at least two nontrivial summands with nonnegative integral components. Note that

$$|\psi_k| = 1, \qquad |k| = 1.$$

Following Siegel [20] we decompose the problem of bounding (9) into two, a simple one involving no divisors, and a not so simple one, involving the small divisors. We define inductively

$$\sigma_r = \sum_{\substack{r_1+\cdots+r_\nu=r \\ \nu\geq 2}} \sigma_{r_1}\cdot\ldots\cdot\sigma_{r_\nu}, \qquad r\geq 2,$$

and

$$\delta_k = \varepsilon_k^{-1} \max_{\substack{k_1+\cdots+k_\nu=k \\ \nu\geq 2}} \delta_{k_1}\cdot\ldots\cdot\delta_{k_\nu}, \qquad |k|\geq 2, \tag{10}$$

with $\sigma_1 = 1$ and $\delta_e = 1$, where e stands for any integer vector k with $|k| = 1$. Then we have

$$|\psi_k| \leq \sigma_{|k|}\delta_k M^{|k|-1}, \qquad |k|\geq 1,$$

as one easily shows by induction. To establish (9) it therefore suffices to prove analogous estimates for the σ_r and δ_k.

Consider first the σ_r. For $\sigma(t) = \sum_{r\geq 1}\sigma_r t^r$ we have

$$\sigma - t = \sum_{r\geq 2}\sigma_r t^r = \sum_{r\geq 2}\left(\sum_{s\geq 1}\sigma_s t^s\right)^r = \frac{\sigma^2}{1-\sigma},$$

hence

$$(\sigma - t)(1-\sigma) = \sigma^2.$$

This equation has a unique analytic solution $\sigma = t + \ldots$, namely

$$\sigma(t) = \frac{t+1}{4}\left(1 - \sqrt{1 - \frac{8t}{(1+t)^2}}\right), \qquad |t| < 1.$$

Hence,

$$\sup_{r\geq 2}\frac{1}{r}\log\sigma_r < \infty.$$

We now consider the δ_k. Here we essentially repeat Brjuno's argument [6]. In (10) the maximum is attained for some decomposition $k = k_1 + \ldots + k_\nu$, which we may choose in some definite way. Decomposing the $\delta_{k_1}, \dots, \delta_{k_\nu}$ in the same manner and proceeding like this, we obtain some well defined decomposition

$$\delta_k = \varepsilon_k^{-1} \varepsilon_{l_1}^{-1} \cdot \dots \cdot \varepsilon_{l_s}^{-1}, \qquad |k| \geq 2,$$

where $2 \leq |l_1|, \dots, |l_s| < |k|$. Moreover,

$$\varepsilon_k = \left|\lambda^k - \lambda_{i_k}\right|, \qquad |k| \geq 2,$$

the index i_k also being chosen in some definite way.

We can then define

$$N_m^j(k), \qquad m \geq 2, \quad 1 \leq j \leq n,$$

to be the number of factors ε_l^{-1} in δ_k, with $l = k, l_1, \dots, l_s$, which satisfy

$$i_l = j \quad \text{and} \quad \varepsilon_l < \theta\omega(m),$$

where

$$4\theta = \min_{1 \leq i \leq n} |\lambda_i| \leq 1. \tag{11}$$

The last inequality can always be satisfied by replacing g by g^{-1} if necessary. Then also $\omega \leq 2$.

The following is the key estimate.

Lemma (Brjuno) *For $m \geq 2$ and $1 \leq j \leq n$,*

$$N_m^j(k) \leq \begin{cases} 0, & |k| \leq m, \\ 2\dfrac{|k|}{m} - 1, & |k| > m. \end{cases}$$

For the proof of the Lemma, we fix m and j, and write N for N_m^j. We then proceed by induction on $|k|$.

For $|k| \leq m$,

$$\varepsilon_k \geq \omega(|k|) \geq \omega(m) \geq \theta\omega(m)$$

by the definition of ω, hence $N(k) = 0$.

So assume that $|k| > m$. Write

$$\delta_k = \varepsilon_k^{-1} \delta_{k_1} \cdot \ldots \cdot \delta_{k_\nu}, \qquad k = k_1 + \cdots + k_\nu, \quad \nu \geq 2,$$

using (10). We distinguish two cases.

Case 1: $\varepsilon_k \geq \theta\omega(m)$ *and* i_k *arbitrary, or* $\varepsilon_k < \theta\omega(m)$ *and* $i_k \neq j$. Then

$$N(k) = N(k_1) + \cdots + N(k_\nu).$$

To each term the induction hypotheses applies, since $|k_1|, \ldots, |k_\nu| < |k|$, from which $N(k) \leq 2\,|k|/m - 1$ follows immediately.

Case 2: $\varepsilon_k < \theta\omega(m)$ *and* $i_k = j$. Then

$$N(k) = 1 + N(k_1) + + \cdots + N(k_\nu).$$

We can assume that $|k| > |k_1| \geq \cdots \geq |k_\nu|$. Again, there are two cases.

Case 2.1: $|k_1| \leq m$ *or* $|k_1| \geq |k_2| > m$. With the first alternative,

$$N(k) = 1 \leq 2\frac{|k|}{m} - 1.$$

With the other alternative, there is $2 \leq \mu \leq \nu$, such that $\left|k_\mu\right| > m \geq \left|k_{\mu+1}\right|$. Then also

$$N(k) = 1 + N(k_1) + \cdots + N(k_\mu) \leq 1 + 2\frac{|k|}{m} - \mu \leq 2\frac{|k|}{m} - 1,$$

and we are done.

Case 2.2: $|k_1| > m \geq |k_2|$. Then

$$N(k) = 1 + N(k_1).$$

Again, we have to distinguish two cases, according to the size of k_1.

Case 2.2.1: $|k_1| \leq |k| - m$. Then

$$N(k) \leq 1 + 2\frac{|k| - m}{m} - 1 \leq 2\frac{|k|}{m} - 1.$$

Case 2.2.2: $|k_1| > |k| - m$. This is the only interesting case. Let $k_* = k_1$. The crucial observation is that $\varepsilon_{k_*}^{-1}$ does not contribute to $N(k_*)$, for if $i_{k_*} = j$, then ε_{k_*} can not be small.

Indeed, suppose $i_{k_*} = j$ and $\varepsilon_{k_*} = \left|\lambda^{k_*} - \lambda_j\right| < \theta\omega(m)$. With (11) we have

$$\left|\lambda^{k_*}\right| > \left|\lambda_j\right| - \theta\omega(m) \geq 4\theta - 2\theta \geq 2\theta.$$

It follows that

$$\begin{aligned}
2\theta\omega(m) &> \varepsilon_k + \varepsilon_{k_*} \\
&= \left|\lambda^k - \lambda_j\right| + \left|\lambda^{k_*} - \lambda_j\right| \\
&\geq \left|\lambda^k - \lambda^{k_*}\right| \\
&\geq \left|\lambda^{k_*}\right| \left|\lambda^{k-k_*} - 1\right| \\
&\geq 2\theta\omega(|k - k_*| + 1) \\
&\geq 2\theta\omega(m),
\end{aligned}$$

which is a contradiction. We applied (2) in the fourth line, because $k - k_*$ has nonnegative integral components, and $1 \leq |k - k_*| = |k| - |k_*| < m$.

Therefore, case 1 applies to δ_{k_*}, and we obtain

$$N(k) = 1 + N(k_{*1}) + \cdots + N(k_{*\nu}),$$

where $|k| > |k_*| > |k_{*1}| \geq \cdots \geq |k_{*\nu}|$, and $k_* = k_{*1} + \ldots k_{*\nu}$ with a different ν. We can now repeat the analysis of case 2 for this decomposition, and we are finished, unless we run again into case 2.2.2. In the latter case,

$$N(k) = 1 + N(k_{**})$$

with $|k_{**}| < |k_*| < |k|$, and we can repeat the above argument. This loop, however, can happen at most m times. Finally, we have to run into a different case. This completes the induction, and the proof of the Lemma.

We can now estimate

$$\frac{1}{|k|} \log \delta_k = \sum_{\mu=0}^{s} \frac{1}{|k|} \log \varepsilon_{k_\mu}^{-1}.$$

Let $1 = q_0 < q_1 < \ldots$ be a sequence of integers for which (3) holds. We can assume that $\sum q_\nu^{-1}$ converges. By the Lemma,

$$\operatorname{card}\left\{0 \le \mu \le s : \theta\omega(q_{\nu+1}) \le \varepsilon_{k_\mu} < \theta\omega(q_\nu)\right\} \le N^1_{q_\nu}(k) + \cdots + N^n_{q_\nu}(k) \le 2n\frac{|k|}{q_\nu}$$

for $\nu \ge 1$. This holds also for $\nu = 0$ when the upper bound is dropped, since the number of *all* factors δ_k is bounded by $2|k| - 1$. This follows directly from (10) by induction. Consequently,

$$\frac{1}{|k|}\log\delta_k \le 2n\sum_{\nu\ge 0}\frac{1}{q_\nu}\log\frac{1}{\theta\omega(q_{\nu+1})} < \infty,$$

for (3) clearly remains valid if ω is multiplied by a constant factor. The right hand side is independent of k, so

$$\sup_k \frac{1}{|k|}\log\delta_k < \infty.$$

The Theorem is proven.

6 Addendum

Here is another proof of Brjuno's Lemma which avoids the study of many different cases. We again fix m and j, and write N for N^j_m. Then we proceed by induction on $|k|$.

For $|k| \le m$,

$$\varepsilon_k \ge \omega(|k|) \ge \omega(m) \ge \theta\omega(m),$$

hence $N(k) = 0$.

So assume that $|k| > m$. Write

$$\delta_k = \varepsilon_k^{-1}\delta_{k_1}\cdot\ldots\cdot\delta_{k_\nu}$$

with

$$k_1 + \ldots + k_\nu = k,$$

$$|k| > |k_1| \ge \cdots \ge |k_\nu| \ge 1.$$

In this decomposition, only $|k_1|$ may be greater than $K = \max(|k| - m, m)$. If this is the case, δ_{k_1} is decomposed in the same way. Repeating this step at most $m - 1$ times, we finally obtain the decomposition

$$\delta_k = \varepsilon_k^{-1} \varepsilon_{k_1}^{-1} \cdot \ldots \cdot \varepsilon_{k_\mu}^{-1} \cdot \delta_{l_1} \cdot \ldots \cdot \delta_{l_\nu},$$

where $\mu \geq 0$, $\nu \geq 2$ and

$$k > k_1 > \cdots > k_\mu,$$
$$l_1 + \cdots + l_\nu = k,$$
$$\left|k_\mu\right| > K \geq |l_1| \geq \cdots \geq |l_\nu|.$$

Here, $k > l$ means that $k - l$ has nonnegative components and is not identically zero.

The point is that at most *one* of the ε's can contribute to $N(k)$. This is the content of Siegel's lemma.

Lemma (Siegel) *If $k > l$ and*

$$\varepsilon_k < \theta\omega(m), \quad \varepsilon_l < \theta\omega(m), \quad i_k = i_l,$$

then $|k - l| \geq m$.

The proof of this lemma is simple. The assumption $\varepsilon_l < \theta\omega(m)$ implies

$$\left|\lambda^l\right| > \left|\lambda_{i_l}\right| - \theta\omega(m) \geq 4\theta - 2\theta = 2\theta.$$

It follows that

$$\begin{aligned} 2\theta\omega(m) &> \varepsilon_k + \varepsilon_l \\ &= \left|\lambda^k - \lambda_{i_k}\right| + \left|\lambda^l - \lambda_{i_l}\right| \\ &\geq \left|\lambda^k - \lambda^l\right| \\ &\geq \left|\lambda^l\right| \left|\lambda^{k-l} - 1\right| \\ &\geq 2\theta\omega(|k - l| + 1), \end{aligned}$$

or $\omega(|k - l| + 1) < \omega(m)$, which implies $|k - l| \geq m$ by the monotonicity of ω. This proves Siegel's lemma.

It now follows from Siegel's lemma and the decomposition of δ_k that

$$N(k) \leq 1 + N(l_1) + \cdots + N(l_\nu).$$

Choose $0 \leq \rho \leq \nu$ such that $|l_\rho| > m \geq |l_{\rho+1}|$. By the induction hypotheses, all terms with $|l| \leq m$ vanish, and we get

$$\begin{aligned} N(k) &\leq 1 + N(l_1) + \cdots + N(l_\rho) \\ &\leq 1 + 2\frac{|l_1 + \cdots + l_\rho|}{m} - \rho \\ &\leq \begin{cases} 1, & \rho = 0, \\ 2\dfrac{|k| - m}{m}, & \rho = 1, \\ 2\dfrac{|l_1 + \cdots + l_\rho|}{m} - 1, & \rho \geq 2 \end{cases} \\ &\leq 2\frac{|k|}{m} - 1. \end{aligned}$$

Brjuno's lemma is proven.

References

[1] V. I. Arnold, Proof of a theorem by A. N. Kolmogorov on the invariance of quasiperiodic motions under small perturbations of the Hamiltonian. *Russ. Math. Surv.* **18** (1963), 9–36.

[2] V. I. Arnold, Singularities of smooth mappings. *Russ. Math. Surv.* **23** (1968), 1–43.

[3] V. I. Arnold, *Geometrical Methods in the Theory of Ordinary Differential Equations.* Springer, 1983.

[4] Y. N. Bibikov, *Local Theory of Nonlinear Analytic Ordinary Differential Equations.* Lecture Notes in Mathematics 702, Springer, 1979.

[5] A. D. Brjuno, Convergence of transformations of differential equations to normal form. *Dokl. Akad. Nauk SSSR* **165** (1965), 987–989.

[6] A. D. Brjuno, Analytic form of differential equations. *Trans. Moscow Math. Soc.* **25** (1971), 131–288, and **26** (1972), 199–239.

[7] T. M. Cherry, A singular case of iteration of analytic functions: A contribution to the small divisor problem. In: *Nonlinear Problems of Engineering,* W. F. Ames (Ed.), New York, 1964, 29–50.

[8] H. Cremer, Über die Häufigkeit der Nichtzentren. *Math. Ann.* **115** (1938), 573–580.

[9] A. GRAY, A fixed point theorem for small divisor problems. *J. Diff. Eq.* **18** (1975), 346–365.

[10] R. S. HAMILTON, The inverse function theorem of Nash and Moser. *Bull. Am. Math. Soc.* **7** (New Series) (1982), 65–232.

[11] W. KLINGENBERG, Typoscript, ETH Zürich, 1984.

[12] G. KOENIGS, Recherches sur les équationes fonctionelles. *Ann. l'Ecole norm.* **1** (1884), Supplement.

[13] A. N. KOLMOGOROV, On the preservation of conditionally periodic motions. *Dokl. Akad. Nauk SSSR* **96** (1954), 527–530.

[14] J. MOSER, On invariant curves of area preserving mappings of an annulus. *Nachr. Akad. Wiss. Göttingen, Math. Phys. Kl.* (1962), 1–20.

[15] J. MOSER, Convergent series expansions for quasiperiodic motions. *Math. Ann.* **169** (1967), 136–176.

[16] H. POINCARÉ, Sur les propriétés des fonctions définies par les équations aux différences partielles. *Ouvres 1,* Gauthier-Villars, 1929.

[17] H. RÜSSMANN, Über die Iteration analytischer Funktionen. *J Math. Mech.* **17** (1967), 523–532.

[18] H. RÜSSMANN, On the convergence of power series transformations of analytic mappings near a fixed point into normal form. Preprint, 1977.

[19] E. SCHRÖDER, Über iterierte Funktionen. *Math. Ann.* **3** (1871), 296–322.

[20] C. L. SIEGEL, Iteration of analytic functions. *Ann. Math.* **43** (1942), 607–612.

[21] C. L. SIEGEL, Über die Normalform analytischer Differentialgleichungen in der Nähe einer Gleichgewichtslösung. *Nachr. Akad. Wiss. Göttingen, Math. Phys. Kl.* (1952), 21–30.

[22] C. L. SIEGEL & J. MOSER, *Lectures on Celestial Mechanics.* Springer, 1971.

[23] S. STERNBERG, Infinite Lie groups and the formal aspects of dynamical systems. *J. Math. Mech.* **10** (1961), 451–474.

[24] S. STERNBERG, *Celestial Mechanics, II.* Benjamin, 1969.

[25] E. ZEHNDER, Generalized implicit function theorems with applications to some small divisor problems I. *Comm. Pure Appl. Math.* **28** (1975), 91–140.

[26] E. ZEHNDER, A simple proof of a generalization of a theorem by C. L. Siegel. *Lecture Notes in Math.* **597** (1977), 855–866.

Mathematisches Institut A, Universität Stuttgart, Pfaffenwaldring 57, D-70569 Stuttgart
poschel@mathematik.uni-stuttgart.de