Ergod. Th. & Dynam. Sys. (2002), **22**, 1551–1573
DOI: 10.1017/S0143385702000974
© *2002 Cambridge University Press*
Printed in the United Kingdom

Stability of elliptic fixed points of analytic area-preserving mappings under the Bruno condition

HELMUT RÜSSMANN
Fachbereich Mathematik und Informatik, Universität Mainz, 55099 Mainz, Germany
(e-mail: ruessmann@mathematik.uni-mainz.de)

Abstract. We show the stability of an elliptic fixed point of an analytic area-preserving mapping under the Bruno condition, using Moser's Twist-mapping Theorem. We require that there are non-vanishing Birkhoff constants and prove the convergence of the transformation to the Birkhoff normal form if all Birkhoff constants are zero. The convergence theorem is formulated and proved for arbitrary dimensions.

1. *Introduction*

We consider a real analytic area-preserving mapping $(x, y) \mapsto (x_1, y_1)$ in the x–y-plane near the fixed point $(0, 0)$ of the form

$$\begin{aligned} x_1 &= f(x, y) = x\cos\gamma_0 - y\sin\gamma_0 + \cdots, \\ y_1 &= g(x, y) = x\sin\gamma_0 + y\cos\gamma_0 + \cdots, \quad f_x g_y - f_y g_x = 1 \end{aligned} \tag{1}$$

where f and g are convergent power series in x, y with real coefficients and $\gamma_0/2\pi$ is an irrational real number. Therefore, the origin $(0, 0)$ is an elliptic fixed point and there exists an area-preserving formal change of variables

$$\begin{aligned} x &= \varphi(\xi, \eta) = \xi + \cdots, \quad x_1 = \varphi(\xi_1, \eta_1), \\ y &= \psi(\xi, \eta) = \eta + \cdots, \quad y_1 = \psi(\xi_1, \eta_1), \quad \varphi_\xi \psi_\eta - \varphi_\eta \psi_\xi = 1 \end{aligned} \tag{2}$$

where $\varphi(\xi, \eta) - \xi, \psi(\xi, \eta) - \eta$ are formal power series in ξ, η with real coefficients, beginning with terms of second degree such that the transformed mapping possesses the Birkhoff normal form

$$\begin{aligned} \xi_1 &= \xi\cos\Gamma(\xi, \eta) - \eta\sin\Gamma(\xi, \eta), \\ \eta_1 &= \xi\sin\Gamma(\xi, \eta) + \eta\cos\Gamma(\xi, \eta), \quad \Gamma(\xi, \eta) = \sum_{k=0}^{\infty} \gamma_k (\xi^2 + \eta^2)^k \end{aligned} \tag{3}$$

where $\gamma_1, \gamma_2, \ldots$ are the uniquely determined Birkhoff constants. These facts are described in detail and proved in §23 of the book of Siegel and Moser [**7**].

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

If not all Birkhoff constants vanish, then Moser's Twist-mapping Theorem can be applied such that in any neighbourhood of the fixed point $(0, 0)$ the mapping (1) has analytic invariant simply closed curves surrounding this fixed point. As a consequence this fixed point is stable (see §32–§34 in [**7**]).

If all Birkhoff constants vanish, $\gamma_1 = \gamma_2 = \cdots = 0$, then we have the following theorem.

THEOREM 1.1. *If there is a formal change of variables (2) transforming the given mapping (1) into the linearized mapping*

$$\xi_1 = \xi \cos \gamma_0 - \eta \sin \gamma_0, \quad \eta_1 = \xi \sin \gamma_0 + \eta \cos \gamma_0 \tag{4}$$

and there exists a function $\Omega : \mathbb{N} \to \mathbb{R}$ *with the properties*

$$|\lambda^k - 1| \geq \frac{1}{\Omega(k)}, \quad k = 1, 2, \ldots, \quad \lambda = e^{i\gamma_0}, \tag{5}$$

$$k \leq \Omega(k) \leq \Omega(k+1), \quad k = 1, 2, \ldots, \tag{6}$$

$$\sum_{k=1}^{\infty} \frac{\log \Omega(k)}{k^2} < \infty, \tag{7}$$

then there is also a convergent change of variables (2) from (1) into (4).

A consequence of this theorem is the fact that under the conditions (5), (6), (7), in the case of vanishing Birkhoff constants the mapping (1) also possesses analytic simply closed invariant curves in any neighbourhood of the fixed point $(0, 0)$, with the latter lying in their interior because the concentric circles around $(0, 0)$ are invariant for the transformed mapping (4). Combining the two cases we obtain the following theorem.

THEOREM 1.2. *If the real number* γ_0 *satisfies the conditions (5), (6), (7) with some function* $\Omega : \mathbb{N} \to \mathbb{R}$, *then the mapping (1) possesses analytic simply closed invariant curves in any neighbourhood of the fixed point* $(0, 0)$ *with the latter lying in their interior. As a consequence this fixed point is stable.*

The significance of this theorem lies in the fact that stability of the fixed point $(0, 0)$ is established only by means of the linear part of the mapping (1). The existence of a function Ω satisfying the relations (5), (6), (7) for γ_0 is equivalent to the Bruno condition. This equivalence will be proved in §8.

Theorem 1.2 was formulated in [**5**] with $\Omega(k) = ck^p$ for $p \geq 2$, $c > 0$, but without proving Theorem 1.1. In 1977, I gave a proof in the I.H.E.S.-Preprint [**6**] which is no longer available. Herman cited this preprint in [**3**] and in his talk [**4**] at the ICM in Berlin, 1998. After his talk he said to me I should finally publish that preprint. This is the purpose of this paper.

The proof of Theorem 1.1 consists of two parts. In the first part we construct a convergent but not necessarily area-preserving transformation from (1) into (4). In the second part we show that there is also a convergent area-preserving transformation (2) from (1) into (4). The first part can be formulated independently of the dimension for mappings in m-space, $m \geq 1$.

In the next section we formulate the convergence theorem for m-dimensional mappings, which is proved in §3–§6. In §7 we shall prove Theorem 1.1.

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

2. *Formulation of the convergence theorem*

We consider a mapping

$$x \mapsto f(x) = Ax + F(x) \tag{8}$$

where

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \in \mathbb{K}^{m\times 1}, \quad A : \mathbb{K}^{m\times 1} \to \mathbb{K}^{m\times 1} \text{ linear}, \quad \mathbb{K} = \mathbb{R} \text{ or } \mathbb{C} \tag{9}$$

and F is an m-column of power series beginning with terms of second degree

$$F = \sum_{k=2}^{\infty} F_k, \quad F_k \in \mathcal{P}^{(k)}, \quad k = 2, 3, \dots \tag{10}$$

and with

$$\mathcal{P}^{(k)} := \left\{ x \mapsto P(x) = \sum_{\substack{j_1+\cdots+j_m=k \\ j_1\geq 0,\dots,j_m\geq 0}} {x_1}^{j_1} \cdots {x_m}^{j_m} P_{j_1\cdots j_m} \;\middle|\; P_{j_1\cdots j_m} \in \mathbb{K}^{m\times 1} \right\},$$
$$k = 1, 2, \dots \tag{11}$$

converging in some neighbourhood of $x = 0$. In order to study the behaviour of the given mapping f, one can try to get a simpler form of (8) by introducing new coordinates. Clearly, the optimal situation would be reached by removing all higher-order terms and reducing f simply to its linear part A. This, however, is not possible in general. However, if there exists a formal change of variables reducing f to its linear part, then there is also an analytic change of variables with this property provided some conditions are satisfied. More precisely we have the following.

THEOREM 2.1. *If f converges in some neighbourhood of $x = 0$ and there exists a formal power series transformation*

$$\xi \mapsto x = \varphi(\xi) = \xi + \sum_{k=2}^{\infty} \Phi_k(\xi), \quad \xi, x \in \mathbb{K}^{m\times 1}, \quad \Phi_k \in \mathcal{P}^{(k)}, \quad k = 2, 3, \dots \tag{12}$$

such that f is reduced to its linear part

$$\varphi^{-1} \circ f \circ \varphi = A, \tag{13}$$

then there is also a convergent transformation of the form (12) with the property (13) provided the following conditions are satisfied ($\mathbb{N} = \{1, 2, \dots\}$, $\mathbb{N}_0 = \{0\} \cup \mathbb{N}$)*:*

(1) *A is diagonalizable and not singular;*

(2) *the eigenvalues $\lambda_1, \dots, \lambda_m$ of A satisfy the inequalities*

$$|\lambda_1^{j_1} \cdots \lambda_m^{j_m} - \lambda_l| \geq \frac{1}{\Omega(j_1 + \cdots + j_m)} \quad \text{for all } j_\alpha \in \mathbb{N}_0, \quad \alpha, l = 1, \dots, m$$
$$\text{with } j_1 + \cdots + j_m \geq 2 \quad \text{and} \quad \lambda_1^{j_1} \cdots \lambda_m^{j_m} \neq \lambda_l \tag{14}$$

where $\Omega : \mathbb{N} \to \mathbb{R}$ is a function restricted by the inequalities

$$k \leq \Omega(k) \leq \Omega(k+1) \quad \text{for all } k \in \mathbb{N} \quad \text{and} \quad \sum_{k=1}^{\infty} \frac{\operatorname{Log} \Omega(k)}{k^2} < \infty. \tag{15}$$

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

In the special case

$$\lambda_1^{j_1} \cdots \lambda_m^{j_m} \neq \lambda_l \quad \text{for all } j_\alpha \in \mathbb{N}_0 \quad \text{and} \quad \alpha, l = 1, \ldots, m \quad \text{with } j_1 + \cdots + j_m \geq 2$$

there exists a uniquely determined formal power series transformation (12) with (13), as one can easily check inductively by comparing coefficients in the equation $f \circ \varphi = \varphi \circ A$ being equivalent to (13). This case has a long history going back to Poincaré. For a survey the reader should consult [**3**].

3. *Formal considerations*

In what follows, we derive the formal facts needed for the proof of Theorem 2.1. We do this in a more general way such that the linear part A of the given mapping f is not necessarily diagonalizable.

As in §2 we consider a mapping (8) with the properties (9)–(11), where, in contrast to §2, F is now only a formal power series without any statement on convergence.

According to the well-known concept of normal forms we try to find a formal power series transformation

$$\mathbb{K}^{m\times 1} \ni \xi \mapsto x = \varphi(\xi) = \xi + \Phi(\xi), \quad \Phi = \sum_{k=2}^{\infty} \Phi_k, \quad \Phi_k \in \mathcal{P}^{(k)}$$

(see (11)) such that

$$\xi \mapsto (\varphi^{-1} \circ f \circ \varphi)(\xi) = n(\xi) = A\xi + N(\xi), \quad N = \sum_{k=2}^{\infty} N_k, \quad N_k \in \mathcal{P}^{(k)} \tag{16}$$

and N gets a form as simple as possible. This equality for φ is obviously equivalent to $\varphi \circ n = f \circ \varphi$. Inserting $\varphi = I + \Phi$ with $I = \mathrm{id}_{\mathbb{K}^{m\times 1}}$, $n = A + N$, $f = A + F$ we obtain

$$N + (\Phi \circ A - A \circ \Phi) = F + F \circ (I + \Phi) - F + \Phi \circ A - \Phi \circ (A + N). \tag{17}$$

Here the operator

$$T : \mathcal{P} \to \mathcal{P}, \quad T\Phi := \Phi \circ A - A \circ \Phi \tag{18}$$

is a linear mapping of the linear space

$$\mathcal{P} := \{\Phi = \Phi_2 + \Phi_3 + \cdots \mid \Phi_k \in \mathcal{P}^{(k)},\ k = 2, 3, \ldots\} \tag{19}$$

of all formal power series with coefficients in $\mathbb{K}^{m\times 1}$ and beginning with at least second-order terms into itself, such that we have

$$T|_{\mathcal{P}^{(k)}} : \mathcal{P}^{(k)} \to \mathcal{P}^{(k)}, \quad k = 2, 3, \ldots . \tag{20}$$

Comparing the polynomials of the kth degree in (17), we get

$$N_k + T\Phi_k = F_k + \mathcal{E}_k(\Phi_j, F_j, N_j,\ j < k), \quad k = 2, 3 \ldots \tag{21}$$

with certain expressions $\mathcal{E}_k$ only depending on polynomials of degree $< k$. Observe that $\mathcal{E}_2 = 0$.

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

In order to obtain reasonable results on normal forms we must at least assume that we have a decomposition

$$\mathcal{P} = \mathcal{N} \oplus \mathcal{M}, \tag{22}$$

such that T has a right inverse on $\mathcal{M}$:

$$T \circ T_r^{-1} = \mathrm{id}_{\mathcal{M}}\,. \tag{23}$$

This simply means that we have such a decomposition for every $\mathcal{P}^{(k)}$:

$$\mathcal{P}^{(k)} = \mathcal{N}^{(k)} \oplus \mathcal{M}^{(k)}, \quad T|_{\mathcal{P}^{(k)}} \circ T_r^{-1}|_{\mathcal{M}^{(k)}} = \mathrm{id}_{\mathcal{M}^{(k)}}, \quad k = 2, 3, \ldots\,. \tag{24}$$

Under these assumptions, equations (21) can easily be solved. Denoting the right members in (21) by H_k we get, in view of (24),

$$N_k = \{H_k\}_{\mathcal{N}^{(k)}}, \quad \Phi_k = T_r^{-1}\{H_k\}_{\mathcal{M}^{(k)}} + G_k$$

where G_k is an arbitrary element in the kernel $\mathcal{K}^{(k)}$ of $T|_{\mathcal{P}^k}$:

$$\mathcal{K}^{(k)} := \{G \in \mathcal{P}^{(k)} \mid TG = 0\}, \quad k = 2, 3, \ldots\,. \tag{25}$$

Since for $k = 2$ the expression $\mathcal{E}_2$ in (21) vanishes, $N_2, N_3, \ldots, \Phi_2, \Phi_3, \ldots$ can be uniquely determined by complete induction provided the $G_k \in \mathcal{K}^{(k)}$, $k = 2, 3, \ldots$ are prescribed. So, denoting the kernel of T by

$$\mathcal{K} := \{G_2 + G_3 + \cdots \mid G_k \in \mathcal{K}^{(k)},\ k = 2, 3, \ldots\} = \{G \in \mathcal{P} \mid TG = 0\}, \tag{26}$$

we have proved the following lemma.

LEMMA 3.1. *Assume that (22), (23) are satisfied. Then, given $g = I + G$, $G \in \mathcal{K}$, there is a unique $\varphi = g + \Phi$, $\Phi \in T_r^{-1}\mathcal{M}$ such that $\varphi^{-1} \circ f \circ \varphi = n$, $n - A \in \mathcal{N}$.*

Definition 3.1. Under the assumptions (22) and (23), we call n a normal form of f in (8)–(11) if $n - A \in \mathcal{N}$ and there exists a transformation $\varphi = g + \Phi$, $g - I \in \mathcal{K}$, $\Phi \in T_r^{-1}\mathcal{M}$ with $\varphi^{-1} \circ f \circ \varphi = n$.

LEMMA 3.2. *Under the assumptions (22) and (23), let f in (8)–(11) already be in normal form up to the terms of degree q, $q \geq 2$; that is, we have*

$$f = A + N_2 + \cdots + N_q + F_{q+1} + \cdots, \quad N_k \in \mathcal{N}^{(k)},$$
$$k = 2, \ldots, q, \quad F_j \in \mathcal{P}^{(j)}, \quad j > q.$$

Then there exists a transformation

$$\varphi = I + \Phi_{q+1} + \Phi_{q+2} + \cdots, \quad \Phi_k \in T_r^{-1}\mathcal{M}^{(k)}, \quad k = q+1, q+2, \ldots$$

of f into a complete normal form $n = \varphi^{-1} \circ f \circ \varphi$, $n - A \in \mathcal{N}$.

Proof. Putting $G_2 = \cdots = G_q = \Phi_2 = \cdots = \Phi_q = 0$ and using (21) for $k = q+1, \ldots$, we obtain, as above, uniquely determined $\Phi_k \in T^{-1}\mathcal{M}^{(k)}$, $k = q+1, \ldots$, if we prescribe $G_{q+1} = G_{q+2} = \cdots = 0$ such that $\varphi = I + \Phi$, $\Phi = \sum_{k=q+1}^{\infty} \Phi_k$ transforms f into a normal form agreeing with f up to terms of degree q. □

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

LEMMA 3.3. *The set* $\mathcal{G} := \{g = I + G \mid G \in \mathcal{K}\}$ *is a group under composition, a subgroup of the group* $\{g = I + G \mid G \in \mathcal{P}\}$ *of all power series transformations beginning with the identity* $I = \mathrm{id}_{\mathbb{K}^{m\times 1}}$.

Proof. For arbitrary power series without constant terms

$$g = G_1 + G, \quad h = H_1 + H, \quad G_1, H_1 \in \mathcal{P}^{(1)}, \quad G, H \in \mathcal{P}$$

(see (11), (19)) we obviously have the equation

$$T(g \circ h) = (Tg) \circ h + g \circ (A \circ h + Th) - g \circ (A \circ h) \tag{27}$$

where the linear operator T is defined as above: $g \mapsto Tg = g \circ A - A \circ g$, but here considered from $\mathcal{P}^{(1)} \oplus \mathcal{P} = \{G_1 + G \mid G_1 \in \mathcal{P}^{(1)},\ G \in \mathcal{P}\}$ into itself. Equation (27) leads to the implication

$$Tg = Th = 0 \Rightarrow T(g \circ h) = 0 \quad \text{for all } g, h \in \mathcal{P}^{(1)} \oplus \mathcal{P} \tag{28}$$

which gives, together with $g \in \mathcal{G} \Rightarrow Tg = 0$, the implication $g, h \in \mathcal{G} \Rightarrow g \circ h \in \mathcal{G}$.

For $g = I + G$, $G \in \mathcal{P}$ we get the inverse $g^{-1} = I + H$, $H \in \mathcal{P}$ by comparing the polynomials of kth degree in the equation $H = -G - (G \circ (I + H) - G)$. We obtain

$$H_2 = -G_2, \quad H_k = -\left[G_k + \sum_{j=2}^{k-1} G_j \circ (I + H_2 + \cdots + H_{k-1})\right]_k, \quad k = 3, 4, \ldots .$$

In the case $g = I + G \in \mathcal{G}$, we have $G \in \mathcal{K}$; hence, $H_2 \in \mathcal{K}^{(2)}$. Assume now that $H_j \in \mathcal{K}^{(j)}$ has already been proved for $2 \le j < k$, then $T(I + H_2 + \cdots + H_{k-1}) = 0$ such that $T[G_j \circ (I + H_2 + \cdots + H_{k-1})] = 0$, $2 \le j < k$, using (28), thus $H_k \in \mathcal{K}^{(k)}$. Complete induction yields $H \in \mathcal{K}$ such that the proof ends with $g \in \mathcal{G} \Rightarrow g^{-1} \in \mathcal{G}$. □

Remark 3.1 Until now, we have required no restrictions for the spaces $\mathcal{M}$ and $\mathcal{N}$ with the exception of (22) and (23). It is natural to choose the space $\mathcal{N}$ as small as possible; that is, $\mathcal{M}$ as large as possible. Since (23) implies $\mathcal{M} \subseteq T\mathcal{P}$, $\mathcal{M} = T\mathcal{P}$ is the natural choice. Then a right inverse T_r^{-1} of T on $\mathcal{M} = T\mathcal{P}$ is uniquely determined by fixing $T_r^{-1}e_{jk}$, $j = 1, \ldots, d_k = \dim \mathcal{M}^{(k)}$ in the inverse image of e_{jk} under T, where $e_{1k}, \ldots, e_{d_k k}$ form a basis of $\mathcal{M}^{(k)}$, $k = 2, 3, \ldots$. Since we have by construction the relations $\dim \mathcal{M}^{(k)} = \dim T_r^{-1}\mathcal{M}^{(k)}$ and $\mathcal{K}^{(k)} \cap T_r^{-1}\mathcal{M}^{(k)} = \{0\}$ and by (24) the equation $\dim \mathcal{K}^{(k)} + \dim \mathcal{M}^{(k)} = \dim \mathcal{P}^{(k)}$, we conclude that $\mathcal{P}^{(k)} = \mathcal{K}^{(k)} \oplus T_r^{-1}\mathcal{M}^{(k)}$, $k = 2, 3, \ldots$; hence, adding (22),

$$\mathcal{P} = \mathcal{N} \oplus \mathcal{M} = \mathcal{K} \oplus T_r^{-1}\mathcal{M}, \quad \mathcal{M} = T\mathcal{P}. \tag{29}$$

Before we obtain results under these conditions, let us remark that sometimes the choice $\mathcal{M} \subsetneq T\mathcal{P}$ is reasonable. Namely, it may happen, for example, that no transformation of the given convergent mapping f in (8)–(11) with (29) converges into a normal form. Then it is reasonable to choose $\mathcal{M}$ small enough such that every convergent $f = A + \cdots$ has a convergent transformation $\varphi = I + \cdots$ into a normal form. Kuiper [**2**, Ch. IV] gives an interesting example for this idea in the analogous case of differential equations $\dot{x} = f(x) = Ax + \cdots$. Also the construction of the stable and unstable manifold near a hyperbolic fixed point of an analytic mapping by means of the normal form approach requires $\mathcal{M} \subsetneq T\mathcal{P}$.

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

LEMMA 3.4. *Let the conditions (29) be satisfied. If* $\Phi \in \mathcal{M}$, $g \in \mathcal{G}$, *then* $\Phi \circ g \in \mathcal{M}$.

Proof. $\Phi \in \mathcal{M}$ implies $\Phi = T\Psi$, $\Psi \in \mathcal{P}$ because of (29). Hence, using (27) we obtain $\Phi \circ g = (T\Psi) \circ g = T(\Psi \circ g) \in \mathcal{M}$. □

LEMMA 3.5. *Let the conditions (29) be satisfied. If* $n^\circ = A + \cdots$ *is a normal form of* f *with (8)–(11), such that we have*

$$g^{-1} \circ n^\circ \circ g - A \in \mathcal{N} \quad \text{for all } g \in \mathcal{G}, \tag{30}$$

then $\{g^{-1} \circ n^\circ \circ g \mid g \in \mathcal{G}\}$ *is precisely the set of normal forms of* f. *If* φ° *is a transformation of* f *into* n°, *then* $\varphi^\circ \circ g$ *is a transformation of* f *into* $g^{-1} \circ n^\circ \circ g$. *Let* $\varphi = I + \Phi$, $\Phi \in T_r^{-1}\mathcal{M}$ *be the uniquely determined transformation of* f *into a normal form* n, *say, with* $G = 0$ *according to Lemma 3.1. Then* $\varphi \circ g = g + \Phi \circ g$ *transforms* f *into the normal form* $g^{-1} \circ n \circ g$ *of* f *for all* $g \in \mathcal{G}$.

Proof. Let $n^* = A + \cdots$ be an arbitrary normal form of f and $\varphi^* = I + \cdots$ a transformation of f into n^* and $\varphi^\circ = I + \cdots$ a transformation of f into the normal form n° such that we have $\varphi^{*-1} \circ f \circ \varphi^* = n^*$ and $\varphi^{\circ -1} \circ f \circ \varphi^\circ = n^\circ$. Then we get

$$\begin{aligned} &\hat{\varphi}^{-1} \circ n^\circ \circ \hat{\varphi} = n^*, \quad \hat{\varphi} := \varphi^{\circ -1} \circ \varphi^* = g + \Phi \quad \text{with some } g \in \mathcal{G},\ \Phi \in T_r^{-1}\mathcal{M}, \\ &\tilde{\varphi}^{-1} \circ n^\circ \circ \tilde{\varphi} = \tilde{n}, \quad \tilde{\varphi} := g \qquad\quad = g + 0 \end{aligned}$$

where $\tilde{n}$ is a normal form because of (30). Therefore, we have two transformations $\hat{\varphi}$ and $\tilde{\varphi}$ transforming the mapping n° into a normal form n^* and $\tilde{n}$, respectively, with the same $g \in \mathcal{G}$. It follows from the uniqueness statement in Lemma 3.1 that $\Phi = 0$. As a consequence we obtain $n^* = g^{-1} \circ n^\circ \circ g$ and $\varphi^* = \varphi^\circ \circ g$.

On the other hand an arbitrary $g \in \mathcal{G}$ yields $(\varphi^\circ \circ g)^{-1} \circ f \circ (\varphi^\circ \circ g) = g^{-1} \circ n^\circ \circ g$.

For the last statement we use what we have proved before: $n = \tilde{g}^{-1} \circ n^\circ \circ \tilde{g}$ with some $\tilde{g} \in \mathcal{G}$. Since $\mathcal{G}$ is a group by Lemma 3.3, we find that $g^{-1} \circ n \circ g = (\tilde{g} \circ g)^{-1} \circ n^\circ \circ (\tilde{g} \circ g)$ is a normal form of f for all $g \in \mathcal{G}$. The lemma is proved. □

THEOREM 3.1. *If the conditions (29) are satisfied and a mapping* f *with (8)–(11) possesses the normal form* $n = A$, *then* A *is the only normal form of* f.

Proof. For $g \in \mathcal{G}$ we have $g \circ A - A \circ g = Tg = 0$, hence $g^{-1} \circ A \circ g - A = 0 \in \mathcal{N}$. Therefore, according to Lemma 3.5, the set of all normal forms of f is $\{A\}$. □

4. *A formal iteration process*

In this section we formally construct an iteration process for a zero

$$\varphi = I + \Phi, \quad I = \mathrm{id}_{\mathbb{K}^{m\times 1}}, \quad \Phi \in \mathcal{P} \tag{31}$$

(see (11), (19)) of the functional equation

$$\mathcal{F}(\varphi) := f \circ \varphi - \varphi \circ A = 0 \tag{32}$$

where

$$f = A + F, \quad F \in \mathcal{P} \tag{33}$$

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

is given. We assume that (29) holds, that A is not singular and that A is a normal form of f according to Definition 3.1. Observe that Lemma 3.2 and Theorem 3.1 can be applied. Then we are in a position to carry out the following iteration process where S' denotes the functional matrix of $S \in \mathcal{P}$; that is, $S'(x) := (\partial S/\partial x_1, \ldots, \partial S/\partial x_m)(x)$ and $U = \mathcal{O}_s$, $U(x) = \mathcal{O}_s(x)$ for $U \in \mathcal{P}$ means that U is a power series beginning with terms of degree at least $s \geq 2$:

$$\left.\begin{aligned}
&\varphi_\nu := I + \Phi_\nu, \quad \Phi_\nu \in \mathcal{P}, \quad \mathcal{F}(\varphi_\nu) = \mathcal{O}_{s_\nu}, \quad s_\nu = 2^{q+\nu} + 1, \quad q \geq 1,\\
&H_\nu := \left(\varphi_\nu{}' \circ A\right)^{-1} \mathcal{F}(\varphi_\nu) = \sum_{k=s_\nu}^{\infty} H_{\nu k},\\
&\qquad H_{\nu k} \in \mathcal{M}^{(k)} \text{ for } s_\nu \leq k < s_{\nu+1}, \quad H_{\nu k} \in \mathcal{P}^{(k)} \text{ for } s_{\nu+1} \leq k,\\
&\Delta\Psi_\nu := T_r^{-1} H_\nu^*, \quad H_\nu^* = \sum_{k=s_\nu}^{s_{\nu+1}-1} H_{\nu k} \in \mathcal{M},\\
&\varphi_{\nu+1} := \varphi_\nu \circ (I + \Delta\Psi_\nu) = I + \Phi_{\nu+1}, \quad \nu = 0, 1, \ldots, \quad \varphi_0 = I.
\end{aligned}\right\} \tag{34}$$

We begin the proof of (34) with checking the first line in (34) for $\nu = 0$, $\varphi_0 = I$:

$$\mathcal{F}(\varphi_0) = f - A = \mathcal{O}_{s_0}, \quad s_0 = 2^q + 1. \tag{35}$$

Without loss of generality we may assume that (35) is true for an arbitrarily large natural number q because this can be achieved by using the existing formal solution (31) of (32) according to Theorem 2.1 and cutting away the terms of degree higher than $s_0 - 1$. Then we even obtain a convergent series, namely a polynomial, $\varphi^* = I + \Phi_2 + \cdots + \Phi_{s_0-1}$ such that $\varphi^{*-1} \circ f \circ \varphi^* - A = \mathcal{O}_{s_0}$. Hence, taking

$$\varphi^{*-1} \circ f \circ \varphi^* \tag{36}$$

in (32) instead of f we have satisfied (35). Moreover, (32) continues to possess a formal solution and if we have a convergent solution φ, we also get a convergent solution $\varphi^* \circ \varphi$ for the equation (32) with the original f. Thus, (35) can be assumed without any loss of generality.

For the proof of the νth step of the process (34) we suppress the ν and suppose

$$\mathcal{F}(\varphi) = \mathcal{O}_s, \quad s \geq 2 \tag{37}$$

to be true for $\varphi = \varphi_\nu$ and $s = s_\nu$. Then we look for a better approximation $\varphi_{\nu+1} = \varphi + \Delta\varphi$ to a zero of equation (32). For this purpose we put

$$\varphi + \Delta\varphi = \varphi \circ (I + \Delta\Psi) \tag{38}$$

and obtain

$$\begin{aligned}
\mathcal{F}(\varphi + \Delta\varphi) &= f \circ (\varphi + \Delta\varphi) - (\varphi + \Delta\varphi) \circ A\\
&= f \circ \varphi \circ (I + \Delta\Psi) - \varphi \circ (I + \Delta\Psi) \circ A\\
&= \mathcal{F}(\varphi) \circ (I + \Delta\Psi) + \varphi \circ A \circ (I + \Delta\Psi) - \varphi \circ (I + \Delta\Psi) \circ A.
\end{aligned} \tag{39}$$

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

Let $S, U \in \mathcal{P}$ and $U = \mathcal{O}_s,\ s \geq 2$, then the formal Taylor series

$$S(x + U(x)) = \sum_{|j|\geq 0} \frac{U(x)^j}{j!} \left(\frac{\partial}{\partial x}\right)^j S(x) = S(x) + S'(x)U(x) + \mathcal{O}_{2s}(x)$$

with the notation

$$\left(\frac{\partial}{\partial x}\right)^j := \left(\frac{\partial}{\partial x_1}\right)^{j_1} \cdots \left(\frac{\partial}{\partial x_m}\right)^{j_m}, \quad x^j := {x_1}^{j_1} \cdots {x_m}^{j_m}, \quad j! := j_1! \cdots j_m!,$$
$$|j| := j_1 + \cdots + j_m, \quad j = (j_1, \ldots, j_m) \in {\mathbb{N}_0}^m \tag{40}$$

is well defined and can be written

$$S \circ (I + U) = S + S'U + \mathcal{O}_{2s} \quad \text{for } S, U \in \mathcal{P},\ U = \mathcal{O}_s \tag{41}$$

such that for the three expressions in the last line of (39) in the case $\Delta\Psi = \mathcal{O}_s$ we get

$$\left.\begin{aligned} \mathcal{F}(\varphi) \circ (I + \Delta\Psi) &= \mathcal{F}(\varphi) + \mathcal{F}(\varphi)'\Delta\Psi + \mathcal{O}_{2s}, \\ \varphi \circ A \circ (I + \Delta\Psi) &= \varphi \circ A + (\varphi \circ A)'\Delta\Psi + \mathcal{O}_{2s}, \\ \varphi \circ (I + \Delta\Psi) \circ A &= \varphi \circ A + (\varphi' \circ A)(\Delta\Psi \circ A) + \mathcal{O}_{2s}. \end{aligned}\right\} \tag{42}$$

Observing that, in view of (37), we have $\mathcal{F}(\varphi)'\Delta\Psi = \mathcal{O}_{2s-1}$ and that $(\varphi \circ A)'\Delta\Psi = (\varphi' \circ A)(A \circ \Delta\Psi)$ we obtain, with (39) and (42),

$$\mathcal{F}(\varphi + \Delta\varphi) = \mathcal{F}(\varphi) + (\varphi' \circ A)(A \circ \Delta\Psi - \Delta\Psi \circ A) + \mathcal{O}_{2s-1}. \tag{43}$$

Since we have $\varphi = I + \Phi$ and $\Phi \in \mathcal{P}$, $\varphi' \circ A$ is a matrix power series (see (40))

$$\varphi'(Ax) = \sum_{|j|\geq 0} x^j Q_j = E + \cdots, \quad Q_j \in \mathbb{K}^{m\times m}, \quad E = m \times m - \text{unit matrix}$$

which is invertible, such that we may form the well-defined power series

$$H := (\varphi' \circ A)^{-1}\mathcal{F}(\varphi) \in \mathcal{P}, \quad H = \mathcal{O}_s, \tag{44}$$

where the second equation is a consequence of (37). By means of (18), (19), (29), (44), we have $H = \{H\}_{\mathcal{M}} + \{H\}_{\mathcal{N}}$, $\{H\}_{\mathcal{M}} = \mathcal{O}_s$, $\{H\}_{\mathcal{N}} = \mathcal{O}_s$ and we may define

$$\Delta\Psi := T_r^{-1}\{H\}_{\mathcal{M}} = \mathcal{O}_s \tag{45}$$

such that the condition $\Delta\Psi = \mathcal{O}_s$ required for the derivation of (43) is fulfilled. Then it follows from (43), (44) and (45) that

$$\mathcal{F}(\varphi + \Delta\varphi) = (\varphi' \circ A)\{H\}_{\mathcal{N}} + \mathcal{O}_{2s-1}. \tag{46}$$

In this formula, nothing changes if we cut the series $H = \sum_{k=s}^{\infty} H_k$, $H_k \in \mathcal{P}^{(k)}$ after the terms of degree $2s - 2$ and replace H by

$$H^* := \sum_{k=s}^{2s-2} H_k \tag{47}$$

in (45) and (46); that is, we define

$$\Delta\Psi := T_r^{-1}\{H^*\}_{\mathcal{M}} = \mathcal{O}_s \tag{48}$$

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

instead of (45) and obtain

$$\mathcal{F}(\varphi + \Delta\varphi) = (\varphi' \circ A)\{H^*\}_{\mathcal{N}} + \mathcal{O}_{2s-1} \tag{49}$$

instead of (46). Now we show that $\{H^*\}_{\mathcal{N}} = 0$. Otherwise assume that $\{H^*\}_{\mathcal{N}} = N_t + \mathcal{O}_{t+1}$, $N_t \in \mathcal{N}^{(t)} \backslash \{0\}$ for some t with $s \le t \le 2s-2$. Then from the definition of $\mathcal{F}(\varphi)$ in (32) and from (49) we get

$$\begin{aligned} f \circ (\varphi + \Delta\varphi) &= (\varphi + \Delta\varphi) \circ A + (\varphi' \circ A)N_t + \mathcal{O}_{t+1} \\ &= (\varphi + \Delta\varphi) \circ A + [(\varphi + \Delta\varphi)' \circ A]N_t + \mathcal{O}_{t+1} \\ &= (\varphi + \Delta\varphi) \circ (A + N_t + \mathcal{O}_{t+1}) \end{aligned} \tag{50}$$

where, in the second line, we inferred from (38), (41), (48) that

$$\Delta\varphi = \varphi \circ (I + \Delta\Psi) - \varphi = \varphi' \Delta\Psi + \mathcal{O}_{2s} = \mathcal{O}_s.$$

Hence, $(\Delta\varphi' \circ A)N_t = \mathcal{O}_{s-1+t} = \mathcal{O}_{t+1}$ because of $s \ge 2$ in (37) and, in the third line, we inferred from (41) and from A being not singular that

$$\begin{aligned} S \circ (A + V) &= (S \circ A) \circ (I + A^{-1} \circ V) = S \circ A + (S \circ A)'(A^{-1} \circ V) + \mathcal{O}_{2t} \\ &= S \circ A + (S' \circ A)V + \mathcal{O}_{2t} \quad \text{with } S = \varphi + \Delta\varphi,\ V = N_t + \mathcal{O}_{t+1}. \end{aligned}$$

As a result we may write (50) in the form

$$(\varphi + \Delta\varphi)^{-1} \circ f \circ (\varphi + \Delta\varphi) = A + N_t + \mathcal{O}_{t+1}. \tag{51}$$

which presents a partial normal form. With the help of Lemma 3.2, we can find a transformation $\varphi^* = I + \mathcal{O}_{t+1}$ such that (51) is transformed into a normal form $A + N_t + N_{t+1} + \cdots$; that is, f is transformed by $(\varphi + \Delta\varphi) \circ \varphi^*$ into this normal form with $N_t \ne 0$. This result contradicts the fact that A is the only normal form because of Theorem 3.1. As a consequence we have $\{H^*\}_{\mathcal{N}} = 0$; hence, using (49), we obtain

$$H^* \in \mathcal{M}, \quad \mathcal{F}(\varphi + \Delta\varphi) = \mathcal{O}_{2s-1}. \tag{52}$$

Obviously the νth step of the complete induction proving (34) is given by the formulae (37), (38), (44), (47), (48) and (52), in view of

$$\varphi = \varphi_\nu, \quad \varphi + \Delta\varphi = \varphi_{\nu+1}, \quad s = s_\nu, \quad 2s - 1 = s_{\nu+1}.$$

5. *Change of variables*

In this section we assume that A is diagonalizable and, for the time being, $\mathbb{K} = \mathbb{C}$. If it is useful we identify the linear mapping A from $\mathbb{C}^{m\times 1}$ into itself with its $m \times m$-matrix in the canonical basis $\{e_1, \ldots, e_m\}$ in $\mathbb{C}^{m\times 1}$, where e_l denotes the lth column of the $m \times m$-unit matrix. There exists a non-singular $m \times m$-matrix C such that $C^{-1}AC = \Lambda := \operatorname{diag}(\lambda_1, \ldots, \lambda_m)$, where $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of A. Besides the space $\mathcal{P}$ of power series in $x_1, \ldots, x_m$ with coefficients in $\mathbb{C}^{m\times 1}$ and without constant and linear terms, we introduce the space $\widetilde{\mathcal{P}}$ of all power series in $y_1, \ldots, y_m$ with coefficients in $\mathbb{C}^{m\times 1}$ and without constant and linear terms, where we relate $x = \operatorname{column}(x_1, \ldots, x_m)$

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

and $y = \text{column}(y_1, \ldots, y_m)$ by the linear change of variables $x = Cy$. There is then a natural isomorphism

$$L : \mathcal{P} \longrightarrow \widetilde{\mathcal{P}}, \qquad L(F)(y) = C^{-1}F(Cy), \quad F \in \mathcal{P},$$
$$L^{-1} : \widetilde{\mathcal{P}} \longrightarrow \mathcal{P}, \quad L^{-1}(\widetilde{F})(x) = C\widetilde{F}(C^{-1}x), \quad \widetilde{F} \in \widetilde{\mathcal{P}}.$$

Since T in (18) is a linear mapping of $\mathcal{P}$ into itself, $\widetilde{T} := LTL^{-1}$ is a linear mapping of $\widetilde{\mathcal{P}}$ into itself defined obviously by $\widetilde{T}(\widetilde{F}) = \widetilde{F}\Lambda - \Lambda\widetilde{F}$. The polynomials

$$y \to e_{jl}(y) := {y_1}^{j_1} \cdots {y_m}^{j_m} e_l, \quad j = (j_1, \ldots, j_m), \quad |j| := \sum_{\alpha=0}^{m} j_\alpha \geq 2, \quad l = 1, \ldots, m$$

form a basis of $\widetilde{\mathcal{P}}$ and we get

$$\widetilde{T}e_{jl} = \lambda_{jl}e_{jl}, \quad \lambda_{jl} := {\lambda_1}^{j_1} \cdots {\lambda_m}^{j_m} - \lambda_l \quad \text{for } |j| \geq 2,\ l = 1, \ldots, m \tag{53}$$

so that λ_{jl} and e_{jl} are the eigenvalues and the eigenvectors of $\widetilde{T}$, respectively. The kernel and range of $\widetilde{T}$ are given by

$$\widetilde{\mathcal{K}} = \left\{ \sum_{\lambda_{jl}=0} a_{jl}e_{jl} \,\middle|\, a_{jl} \in \mathbb{C} \right\}, \quad \widetilde{T}\widetilde{\mathcal{P}} = \left\{ \sum_{\lambda_{jl}\neq 0} a_{jl}e_{jl} \,\middle|\, a_{jl} \in \mathbb{C} \right\} \tag{54}$$

so that we get $\widetilde{\mathcal{P}} = \widetilde{\mathcal{K}} \oplus \widetilde{T}\widetilde{\mathcal{P}}$. From $\widetilde{T}L = LT$ we immediately obtain

$$\mathcal{K} = L^{-1}\widetilde{\mathcal{K}}, \quad T\mathcal{P} = L^{-1}\widetilde{T}\widetilde{\mathcal{P}}, \quad \mathcal{P} = \mathcal{K} \oplus T\mathcal{P}.$$

From (53), we see that $\widetilde{T}$ maps $\widetilde{T}\widetilde{\mathcal{P}}$ onto itself and the right inverse on $\widetilde{T}\widetilde{\mathcal{P}}$ is defined by

$$\widetilde{T}_r^{-1}e_{jl} = \frac{1}{\lambda_{jl}}e_{jl}, \quad \lambda_{jl} \neq 0. \tag{55}$$

Correspondingly, $T_r^{-1} = L^{-1}\widetilde{T}_r^{-1}L$ maps $T\mathcal{P}$ onto itself. Choosing

$$\mathcal{N} = \mathcal{K}, \quad \mathcal{M} = T\mathcal{P}, \tag{56}$$

we have the situation (29) because of $\mathcal{M} = T\mathcal{M} = T_r^{-1}\mathcal{M}$. Moreover, the condition (30) of Lemma 3.5 is satisfied. This fact follows easily from Lemma 3.3 and the considerations at the beginning of the proof of this lemma, where the operator T is extended to the space $\mathcal{P}^{(1)} \oplus \mathcal{P}$. Since $TA = 0$, we get $Tn = 0$ for any normal form n, $n - A \in \mathcal{N} = \mathcal{K}$. Hence, we have $T(g^{-1} \circ n \circ g) = 0 \Rightarrow g^{-1} \circ n \circ g - A \in \mathcal{N} = \mathcal{K}$ because of (28) and $g \in \mathcal{G} \Rightarrow g^{-1} \in \mathcal{G}$.

As a result the assertions of Lemma 3.5 hold and also Theorem 3.1 can be applied.

Now we assume that we have a power series $\widetilde{F} = \widetilde{G} + \widetilde{\Phi} \in \widetilde{\mathcal{P}}$, $\widetilde{G} \in \widetilde{\mathcal{K}}$, $\widetilde{\Phi} \in \widetilde{T}\widetilde{\mathcal{P}}$ which converges (absolutely in some neighbourhood of the origin). Then $\widetilde{G}$ and $\widetilde{\Phi}$ also converge. This is clear from the representation of $\widetilde{\mathcal{K}}$ and $\widetilde{T}\widetilde{\mathcal{P}}$ in (54).

LEMMA 5.1. *If $F = G + \Phi$, $G \in \mathcal{K}$, $\Phi \in \mathcal{M}$ and F converges, then G and Φ also converge.*

Proof. We consider $LF = LG + L\Phi$, $LG \in \widetilde{\mathcal{K}}$, $L\Phi \in \widetilde{T}\widetilde{\mathcal{P}}$. Since F converges, $y \mapsto LF(y) = C^{-1}F(Cy)$ also converges. So LG and $L\Phi$ converge according to the remark just made. Hence, G and Φ converge. □

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

LEMMA 5.2. *If $g + \Phi \circ g$ with $g \in \mathcal{G}$, $\Phi \in \mathcal{M}$ converges, then g and Φ also converge.*

Proof. Lemma 3.4 implies $\Phi \circ g \in \mathcal{M}$. Therefore we have convergence of g and $\Phi \circ g$ by Lemma 5.1. However, then g^{-1} converges and so $\Phi = (\Phi \circ g) \circ g^{-1}$ is convergent. □

LEMMA 5.3. *If there is a convergent transformation of the given mapping $f = A + \cdots$ into a normal form, then the uniquely-defined transformation $I + \Phi$, $\Phi = \Phi_I \in \mathcal{M}$ also converges into a normal form.*

Proof. Since by Lemma 3.5 all transformations into a normal form are given by $g + \Phi \circ g$, $g \in \mathcal{G}$, we have by Lemma 5.2 convergence of Φ and so of $I + \Phi$. □

Now let us assume that the conditions of Theorem 2.1 are fulfilled for our given mapping $f = A + F$, $\mathbb{K} = \mathbb{C}$. We have to find a convergent solution $\varphi = I + \Phi$, $\Phi \in \mathcal{P}$ of the equation

$$f \circ \varphi = \varphi \circ A \tag{57}$$

where a formal solution is known. Changing variables in the above-mentioned way, we have to solve

$$\tilde{f} \circ \tilde{\varphi} = \tilde{\varphi} \circ \Lambda, \quad \tilde{f} = \Lambda + LF, \quad \tilde{\varphi} = I + L\Phi, \tag{58}$$

that is, the same problem with $A = \Lambda$. Having solved this latter problem we get a convergent solution of (58) and Lemma 5.3 gives us the convergence of the 'canonical' solution.

Finally, let us turn to the case $\mathbb{K} = \mathbb{R}$. This means that A is real and F has real coefficients. Solving (58) we obtain in the described way ($\mathbb{K} = \mathbb{C}$) the convergence of the uniquely determined solution

$$\varphi = I + \Phi, \quad \Phi \in \mathcal{M} \tag{59}$$

of (57) with possible complex coefficients. Taking the complex conjugate in (57), we obtain $f \circ \overline{\varphi} = \overline{\varphi} \circ A$ with $\overline{\varphi} = I + \overline{\Phi}$. Since $\Phi \in \mathcal{M}$ there is some Ψ with $\Phi = T\Psi = \Psi \circ A - A \circ \Psi$; hence, we have $\overline{\Phi} = \overline{T\Psi} = T\overline{\Psi} \in \mathcal{M}$ and by the uniqueness of (59) finally $\Phi = \overline{\Phi}$. So we have reduced the proof of Theorem 2.1 to the case $\mathbb{K} = \mathbb{C}$, $A = \Lambda$.

6. *Proof of convergence*

In this section we prove the existence of a convergent solution $\varphi = I + \Phi$, $\Phi \in \mathcal{P}$ of the equation

$$\mathcal{F}(\varphi) := f \circ \varphi - \varphi \circ A = 0 \tag{60}$$

where

$$f = A + F, \quad F \in \mathcal{P} \tag{61}$$

satisfies the conditions of Theorem 2.1. According to the remark at the end of §5, we may assume that $\mathbb{K} = \mathbb{C}$ and A has diagonal form $A = \Lambda$. In this case, we have $\mathcal{P} = \mathcal{K} \oplus T\mathcal{P}$ and we may choose $\mathcal{N} = \mathcal{K}$ and $\mathcal{M} = T\mathcal{P}$ as described in §5 such that (29) holds.

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

Therefore, we may apply the iteration process (34). For proving the convergence of this process we use the following definitions:

$$\left.\begin{aligned} |x| &:= \max_{1\le j\le m} |x_j| \quad \text{for all} \quad x = \operatorname{column}(x_1,\dots,x_m) \in \mathbb{C}^{m\times 1},\\ |C| &:= \sup_{|x|\le 1} |Cx| \quad \text{for all} \quad C \in \mathbb{C}^{m\times m},\\ \mathcal{B}_\varrho &:= \{x \in \mathbb{C}^{m\times 1} \mid |x| < \varrho\}, \qquad A\mathcal{B}_\varrho := \{Ax \mid x \in \mathcal{B}_\varrho\},\\ \mathcal{D}_\varrho &:= \mathcal{B}_\varrho \cup A\mathcal{B}_\varrho \quad \text{for all} \quad \varrho > 0,\\ |F|_{\mathcal{D}} &:= \sup_{x\in\mathcal{D}} |F(x)| \quad \text{for all} \quad F : \mathbb{C}^{m\times 1} \to \mathbb{C}^{m\times 1}, \quad \mathcal{D} \subseteq \mathbb{C}^{m\times 1}. \end{aligned}\right\} \tag{62}$$

According to the conditions of Theorem 2.1, the mapping f in (61) is holomorphic in some open neighbourhood of the origin $0 \in \mathbb{C}^{m\times 1}$. We assume that f has already been prepared for the application of the iteration process (34) such that (35) holds for a sufficiently large positive integer q to be determined later.

Without loss of generality we may assume that f in (61) is holomorphic in the ball $\mathcal{B}_{1+|A|}$ and satisfies the inequality

$$|f - A|_{\mathcal{B}_{2+|A|}} \le 1. \tag{63}$$

This reduction is carried out by means of the change of variables $x = \varepsilon\hat{x}$ with $\varepsilon > 0$ sufficiently small. Then φ and f in (60) have to be replaced by

$$\hat{x} \mapsto \frac{1}{\varepsilon}\varphi(\varepsilon\hat{x}) \quad \text{and} \quad \hat{x} \mapsto \frac{1}{\varepsilon} f(\varepsilon\hat{x}) = A + \varepsilon^{2^q}\mathcal{O}_{s_0}$$

in view of (35). This way the domain of holomorphy of f can be made arbitrarily large and simultaneously $f - A$ can be made arbitrarily small in a fixed ball so that (63) can be satisfied.

The scheme of estimates for the iteration process (34) now runs as follows:

$$\left.\begin{aligned} \mathcal{F}(\varphi_\nu) = \mathcal{O}_{s_\nu}, \quad s_\nu = 2^{q+\nu} + 1,\\ |\varphi_\nu|_{\mathcal{D}_{\varrho_\nu}} \le 1 + |A|,\\ |\varphi_\nu' - E|_{\mathcal{D}_{\varrho_\nu}} \le \sum_{\mu=0}^{\nu-1} \varepsilon_\mu \le \tfrac{1}{2},\\ \varrho_{\nu+1} = \varrho_\nu - 2\delta_\nu, \quad \nu = 0,1,\dots, \quad \varrho_0 = 1 \end{aligned}\right\} \tag{64}$$

where E is the $m \times m$-unit matrix and $\delta_\nu, \varepsilon_\nu, \varrho_\nu, \nu = 0, 1, \dots$ are positive constants.

The sequence $\varrho_0, \varrho_1, \dots$ is completely determined: $\varrho_\nu = 1 - 2\sum_{\mu=0}^{\nu-1}\delta_\mu$. Thus, in order to have a common open domain of definition $\mathcal{D}_{\varrho_\infty} \subseteq \mathcal{D}_{\varrho_\nu}$, $\nu = 0, 1, \dots$ we must have $\varrho_\nu \ge \varrho_\infty > 0$ or, say,

$$\sum_{\mu=0}^{\infty} \delta_\mu \le \tfrac{1}{3}. \tag{65}$$

Furthermore, from (64) we get

$$\sum_{\mu=0}^{\infty} \varepsilon_\mu \le \tfrac{1}{2} \tag{66}$$

as a necessary condition.

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

Now let us assume that we are in a position to determine sequences of positive numbers $\delta_0, \delta_1, \ldots$, and $\varepsilon_0, \varepsilon_1, \ldots$ in such a way that (64) is valid for the process (34) and that (65) and (66) are satisfied. Then the sequence of (non-homogeneous) polynomials $\varphi_0 = I, \varphi_1 = I + \cdots, \ldots$ is uniformly bounded in $\mathcal{D}_{1/3}$. Therefore, we can select a subsequence $\varphi_{\nu_1} = I + \cdots, \varphi_{\nu_2} = I + \cdots, \ldots$ which converges uniformly in $\mathcal{D}_{1/3}$ towards a holomorphic function φ_∞ in $\mathcal{D}_{1/3}$, the Taylor expansion of which at the origin begins with the identity $\varphi_\infty = I + \cdots$. Moreover, from (62), (63), (64), we infer that the sequence $\mathcal{F}(\varphi_{\nu_1}), \mathcal{F}(\varphi_{\nu_2}), \ldots$ of holomorphic functions on $\mathcal{B}_{1/3}$ converges towards the holomorphic function $\mathcal{F}(\varphi_\infty)$ on $\mathcal{B}_{1/3}$ the Taylor series at the origin of which has no non-vanishing coefficient. As a consequence $\varphi = \varphi_\infty = I + \cdots$ is the convergent solution of (60) we want.

Now we turn to the proof of the estimates in (64) by complete induction. For $\varphi_0 = I$ they are valid because of

$$|\varphi_0|_{\mathcal{D}_{\varrho_0}} = |I|_{\mathcal{D}_1} \le |I|_{\mathcal{B}_1} + |I|_{A\mathcal{B}_1} \le 1 + |A|, \quad |\varphi_0{}' - E|_{\mathcal{D}_1} = |E - E|_{\mathcal{D}_1} = 0.$$

So let us assume that the inequalities in (64) are valid for some $\nu \ge 0$ and prove that they are valid for $\nu + 1$. In the course of this proof we suppress the index ν and write $\varrho, s, H, \Delta\Phi, \ldots$ instead of $\varrho_\nu, s_\nu, H_\nu, \Delta\Phi_\nu, \ldots$.

Since by assumption we have $|\varphi(x)| \le 1 + |A|$ for $x \in \mathcal{D}_\varrho$, $\mathcal{F}(\varphi)$ is a holomorphic function defined in $\mathcal{B}_\varrho$, in view of the definition in (60), of (62), and (63), and we obtain the estimates

$$\begin{aligned} |\mathcal{F}(\varphi)(x)| &\le |f(\varphi(x))| + |\varphi(A(x))| \le |(f - A)(\varphi(x))| + |A\varphi(x)| + |\varphi(Ax)| \\ &\le 1 + |A||\varphi(x)| + |\varphi(Ax)| \le 1 + |A|(1 + |A|) + 1 + |A|, \quad |x| < \varrho, \end{aligned}$$

hence,

$$|\mathcal{F}(\varphi)|_{\mathcal{B}_\varrho} \le c_1 := 1 + (1 + |A|)^2. \tag{67}$$

According to (34) we must estimate $H = (\varphi' \circ A)^{-1}\mathcal{F}(\varphi)$. Since by (64) we have $|\varphi'(x) - E| \le 2^{-1}$ for $x \in \mathcal{D}_\varrho$, we get for the inverse of the functional matrix $\varphi'(x)$ in the usual way

$$|\varphi'(x)^{-1}| = |[E - (\varphi'(x) - E)]^{-1}| \le \sum_{j=0}^{\infty} |\varphi'(x) - E|^j \le \sum_{j=0}^{\infty} 2^{-j} = 2, \quad x \in \mathcal{D}_\varrho,$$

thus, $|\varphi'(Ax)^{-1}| \le 2$, $x \in \mathcal{B}_\varrho$. This inequality leads with (67) to the estimate $|H(x)| \le |\varphi'(Ax)^{-1}||\mathcal{F}(\varphi)(x)| \le 2c_1$, $x \in \mathcal{B}_\varrho$ or

$$|H|_{\mathcal{B}_\varrho} \le 2c_1. \tag{68}$$

Following (34) we have to use the expansion

$$H = \sum_{k=s}^{\infty} H_k, \quad H_k \in \mathcal{P}^{(k)}, \quad H^* = \sum_{k=s}^{2s-2} H_k \in \mathcal{M} \quad \text{and} \quad \Delta\Psi = T_r^{-1} H^*. \tag{69}$$

In order to estimate these functions we use the considerations in §5. Recall that at the beginning of the present section we have assumed that $A = \Lambda$, so that we may put $C = E$,

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

$x = y$ and remove $\sim$ from all symbols in §5. With the canonical basis $\{e_{jl}\}$ we can then express (69) in the form

$$H = \sum_{\substack{s\le|j|\\1\le l\le m}} H_{jl}e_{jl}, \quad H^* = \sum{}' H_{jl}e_{jl}, \quad \text{where } \sum{}' = \sum_{\substack{\lambda_{jl}\neq 0\\ s\le|j|\le 2s-2\\ 1\le l\le m}}, \tag{70}$$

$$\Delta\Psi = \sum{}' \frac{1}{\lambda_{jl}} H_{jl}e_{jl}, \quad \lambda_{jl} = \lambda_1{}^{j_1}\cdots\lambda_m{}^{j_m} - \lambda_l \tag{71}$$

where we have still observed (53) and (55). The first equation here leads to $H_l(x) = \sum_{s\le|j|} H_{jl}x^j$ for the lth component H_l of H. Since we get $|H_l|_{\mathcal{B}_\varrho} \le 2c_1$ from (68), Cauchy's estimate for the coefficients of power series yields

$$|H_{jl}| \le \frac{2c_1}{\varrho^{|j|}}, \quad |j| \ge s, \quad l = 1,\dots,m. \tag{72}$$

Therefore, we can estimate H^* in (70), using $e_{jl}(x) = x^j e_l$, $|e_l| = 1$:

$$\begin{aligned} |H^*(x)| &\le \sum{}' |H_{jl}||e_{jl}(x)| \le 2mc_1 \sum{}' \left(1-\frac{\delta}{\varrho}\right)^{|j|} \\ &\le 2mc_1\left(1-\frac{\delta}{\varrho}\right)^s \sum_{|j|\ge 0}\left(1-\frac{\delta}{\varrho}\right)^{|j|} \\ &= 2mc_1\left(1-\frac{\delta}{\varrho}\right)^s \left(\frac{\varrho}{\delta}\right)^m, \quad |x| < \varrho-\delta \end{aligned}$$

such that we obtain

$$|H^*|_{\mathcal{B}_{\varrho-\delta}} \le 2mc_1 e^{-s\delta}\delta^{-m}, \quad 0<\delta<\varrho\le 1 \tag{73}$$

because we have $\varrho = \varrho_\nu \le 1$ and $1-\delta/\varrho \le 1-\delta \le e^{-\delta}$ for $0<\delta<\varrho\le 1$. Similarly, from (14), (15), (71) and (72) we get

$$\begin{aligned} |\Delta\Psi(x)| &\le \sum{}' \frac{1}{|\lambda_{jl}|}|H_{jl}||e_{jl}(x)| \le 2mc_1\Omega(2s-2)\left(1-\frac{\delta}{\varrho}\right)^s \sum_{|j|\ge 0}\left(1-\frac{\delta}{\varrho}\right)^{|j|} \\ &= 2mc_1\Omega(2s-2)\left(1-\frac{\delta}{\varrho}\right)^s\left(\frac{\varrho}{\delta}\right)^m, \quad |x|<\varrho-\delta \end{aligned}$$

and finally obtain

$$|\Delta\Psi|_{\mathcal{B}_{\varrho-\delta}} \le 2mc_1\Omega(2s-2)e^{-s\delta}\delta^{-m}, \quad 0<\delta<\varrho\le 1. \tag{74}$$

For the estimates to follow we need the next lemma.

LEMMA 6.1. *We have $x + t\Delta\Psi(x) \in \mathcal{D}_\varrho$ for $x \in \mathcal{D}_{\varrho-\delta}$, $0\le t\le 1$, $0<\delta<\varrho\le 1$ under the condition*

$$c_2\Omega(2s-2)e^{-s\delta}\delta^{-m} \le \frac{\delta}{2}, \quad c_2 := 2mc_1(1+|A^{-1}|). \tag{75}$$

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

Proof. First we note that (75) implies

$$|\Delta\Psi(x)| \le \frac{\delta}{2}, \quad |A^{-1}||H^*(x)| \le \frac{\delta}{2} \quad \text{for all } x \in \mathcal{B}_{\varrho-\delta}, \tag{76}$$

in view of (73), (74) and $\Omega(k) \ge k \ge 1$ in (15). A^{-1} exists because of condition 1 in Theorem 2.1. Recalling that $\Delta\Psi = T_r^{-1}H^*$ in (69), we have

$$T\Delta\Psi = \Delta\Psi \circ A - A \circ \Delta\Psi = H^*, \tag{77}$$

hence,

$$A + t\Delta\Psi \circ A = A \circ (I + t\Delta\Psi + tA^{-1}H^*), \quad 0 \le t \le 1 \tag{78}$$

and with (76) the implications $x \in A\mathcal{B}_{\varrho-\delta} \Longrightarrow x = Ay$ for $y \in \mathcal{B}_{\varrho-\delta} \Longrightarrow |y + t\Delta\Psi(y) + tA^{-1}H^*(y)| \le |y| + |\Delta\Psi(y)| + |A^{-1}||H^*(y)| \le \varrho - \delta + (\delta/2) + (\delta/2) = \varrho$, such that (78) applied to y leads to

$$x + t\Delta\Psi(x) \in A\mathcal{B}_\varrho \quad \text{for all } x \in A\mathcal{B}_{\varrho-\delta}, \quad 0 \le t \le 1. \tag{79}$$

On the other hand, we have $|x + t\Delta\Psi(x)| \le |x| + |\Delta\Psi(x)| < \varrho - \delta + (\delta/2) < \varrho$ for all $x \in \mathcal{B}_{\varrho-\delta}$, $0 \le t \le 1$, in view of (76); hence, $x + t\Delta\Psi(x) \in \mathcal{B}_\varrho$ for all $x \in \mathcal{B}_{\varrho-\delta}$, $0 \le t \le 1$. This result together with (79) yields the assertion of the lemma because of the definition of $\mathcal{D}_\varrho$ in (62) which is very useful in this context and was first introduced by Zehnder [**8**]. □

In order to extend the estimate (74) to the domain $\mathcal{D}_{\varrho-\delta}$ we use equation (77) a second time and get, together with (73), (74) and $\Omega(k) \ge k \ge 1$ in (15), $|\Delta\Psi(Ax)| \le |A\Delta\Psi(x)| + |H^*(x)| \le c_3\Omega(2s-2)e^{-s\delta}\delta^{-m}$ for $x \in \mathcal{B}_{\varrho-\delta}$ and so

$$|\Delta\Psi|_{\mathcal{D}_{\varrho-\delta}} \le c_3\Omega(2s-2)e^{-s\delta}\delta^{-m}, \quad 0 < \delta < \varrho \le 1, \quad c_3 := 2mc_1(1+|A|). \tag{80}$$

Supposing now the condition (75) to be satisfied we are in a position to estimate the functional matrix of

$$\Delta\varphi := \varphi \circ (I + \Delta\Psi) - \varphi. \tag{81}$$

Observe that from (64) we have $|\varphi'(x)| \le |\varphi'(x) - E| + |E| \le (1/2) + 1 = 3/2$ for $x \in \mathcal{D}_\varrho$. Thus, we obtain, using Lemma 6.1 and (81),

$$\begin{aligned} |\Delta\varphi(x)| &= \left| \int_0^1 \frac{d}{dt}\varphi(x + t\Delta\Psi(x))\, dt \right| = \left| \int_0^1 \varphi'(x + t\Delta\Psi(x))\, dt\Delta\Psi(x) \right| \\ &\le \int_0^1 |\varphi'(x + t\Delta\Psi(x))|\, dt|\Delta\Psi(x)| \le \frac{3}{2}|\Delta\Psi(x)| \le \frac{3}{2}|\Delta\Psi|_{\mathcal{D}_{\varrho-\delta}}, \quad x \in \mathcal{D}_{\varrho-\delta}. \end{aligned}$$

From these results it follows that the functions $t \mapsto \Delta\varphi(x+ta)$, $t \mapsto (\Delta\varphi \circ A)(x+ta)$ are defined and are, therefore, holomorphic for arbitrary fixed $x \in \mathcal{B}_{\varrho-2\delta}$, $a \in \mathcal{B}_1$ with $0 < \delta < (\varrho/2)$ in the ball $\{t \in \mathbb{C} \mid |t| < \delta\}$ and can be estimated by $(3/2)|\Delta\Psi|_{\mathcal{D}_{\varrho-\delta}}$. Consequently the derivative of these functions at $t = 0$ can be estimated by $(3/2\delta)|\Delta\Psi|_{\mathcal{D}_{\varrho-\delta}}$, using Cauchy's formula. Thus, we get

$$|\Delta\varphi'(x)a|, |\Delta\varphi'(Ax)Aa| \le \frac{3}{2\delta}|\Delta\Psi|_{\mathcal{D}_{\varrho-\delta}}, \quad x \in \mathcal{B}_{\varrho-2\delta}, \quad a \in \mathcal{B}_1.$$

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

Hence, $|\Delta\varphi'(x)| \le (3/2\delta)|\Delta\Psi|_{\mathcal{D}_{\varrho-\delta}}$, $x \in \mathcal{B}_{\varrho-2\delta}$ and $|\Delta\varphi'(Ax)| = |\Delta\varphi'(Ax)AA^{-1}| \le |\Delta\varphi'(Ax)A||A^{-1}| \le (3/2\delta)|A^{-1}||\Delta\Psi|_{\mathcal{D}_{\varrho-\delta}}$, $x \in \mathcal{B}_{\varrho-2\delta}$. Finally we obtain $|\Delta\varphi'(x)| \le (3/2\delta)(1+|A^{-1}|)|\Delta\Psi|_{\varrho-\delta}$ for all $x \in \mathcal{D}_{\varrho-2\delta}$ or, with (80),

$$|\Delta\varphi'|_{\mathcal{D}_{\varrho-2\delta}} \le c_4\Omega(2s-2)e^{-s\delta}\delta^{-m-1}, \quad 0 < 2\delta < \varrho \le 1, \quad c_4 := \tfrac{3}{2}(1+|A^{-1}|)c_3. \tag{82}$$

Now we provide all the relevant symbols with the index ν and observe the last line of (64). Then the inequality for $\Delta\varphi'$ in (82) reads

$$|\Delta\varphi_\nu'|_{\mathcal{D}_{\varrho_{\nu+1}}} \le \varepsilon_\nu := c_4\Omega(s_{\nu+1}-1)e^{-s_\nu\delta_\nu}{\delta_\nu}^{-m-1}. \tag{83}$$

Since we have $\varphi_{\nu+1} = \varphi_\nu + \Delta\varphi_\nu = \varphi_\nu \circ (I + \Delta\Psi_\nu)$ by (34) and (81) and $\mathcal{D}_{\varrho_{\nu+1}} \subseteq \mathcal{D}_{\varrho_\nu}$, we get, with (64),

$$|\varphi_{\nu+1}' - E|_{\mathcal{D}_{\varrho_{\nu+1}}} \le |\varphi_\nu' - E|_{\mathcal{D}_{\varrho_\nu}} + |\Delta\varphi_\nu'|_{\mathcal{D}_{\varrho_{\nu+1}}} \le \sum_{\mu=0}^{\nu-1}\varepsilon_\mu + \varepsilon_\nu = \sum_{\mu=0}^{\nu}\varepsilon_\mu.$$

Moreover, $|\varphi_{\nu+1}|_{\mathcal{D}_{\varrho_{\nu+1}}} \le 1 + |A|$, because Lemma 6.1 yields $(I + \Delta\Psi_\nu)(x) \in \mathcal{D}_{\varrho_\nu}$ for $x \in \mathcal{D}_{\varrho_{\nu+1}}$. Since, according to (64), the inequality $\varepsilon_\nu \le 1/2$ for ϵ_ν in (83) must be satisfied and we have, in view of (67), (75), (80) and (82),

$$c_4 = \tfrac{3}{2}(1+|A|)c_3 > c_2 > c_1 > 1, \tag{84}$$

the condition (75) of Lemma 6.1 is fulfilled.

As a result, the estimates in (64) are valid for $\nu+1$ instead of ν and we are left with the problem of finding a sequence of positive constants $\delta_0, \delta_1, \ldots$, such that the inequalities (65) and (66) are satisfied where ε_ν, $\nu = 0, 1, \ldots$, are defined in (83). Since we must have $\delta_\nu \le 1$, $\varepsilon_\nu \le 1$, in view of (65) and (66), we get from (83) and (84) the necessary condition $\Omega(s_{\nu+1}-1)e^{-s_\nu\delta_\nu} \le 1$ or $\delta_\nu \ge {s_\nu}^{-1}\operatorname{Log}\Omega(s_{\nu+1}-1)$. Therefore, we define δ_ν as is done in Lemma 6.2 below. According to this lemma the inequalities (65) and (66) are satisfied if q is chosen so large that

$$4(m+2)\sum_{k=2^{q+1}}^{\infty}\frac{\operatorname{Log}\Omega(k)}{k^2} \le \frac{1}{3}, \quad \frac{c_4}{q} \le \frac{1}{2}, \quad c_4 = 3m[1+(1+|A|)^2](1+|A|)(1+|A^{-1}|).$$

This is clearly possible in the dependence of $|A|, |A^{-1}|, m, \Omega$. Theorem 2.1 is proved.

LEMMA 6.2. *We define* $s_\nu := 2^{q+\nu}+1$,

$$\delta_\nu := (m+2)s_\nu^{-1}\operatorname{Log}\Omega(s_{\nu+1}-1), \quad \eta_\nu := \Omega(s_{\nu+1}-1)\delta_\nu^{-m-1}e^{-s_\nu\delta_\nu}$$

for $\nu = 0, 1, \ldots$, *with* $m, q \in \mathbb{N}$ *and* Ω *satisfying (15). Then we have the estimates*

$$\sum_{\nu=0}^{\infty}\delta_\nu \le 4(m+2)\sum_{k=2^{q+1}}^{\infty}\frac{\operatorname{Log}\Omega(k)}{k^2}, \quad \sum_{\nu=0}^{\infty}\eta_\nu \le \frac{1}{q}.$$

Proof. For integers a, b with $0 < a < b$ we have

$$\frac{1}{a} - \frac{1}{b} = \sum_{k=a}^{b-1}\left\{\frac{1}{k} - \frac{1}{k+1}\right\} = \sum_{k=a}^{b-1}\frac{1}{k(k+1)} \le \sum_{k=a}^{b-1}\frac{1}{k^2},$$

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

hence,

$$\frac{1}{2^{p+1}} = \frac{1}{2^p} - \frac{1}{2^{p+1}} \le \sum_{k=2^p}^{2^{p+1}-1} \frac{1}{k^2}, \quad p = 0, 1, \dots .$$

Since Ω in (15) is not decreasing, we obtain

$$\frac{\operatorname{Log}\Omega(2^p)}{2^{p+1}} \le \sum_{k=2^p}^{2^{p+1}-1} \frac{\operatorname{Log}\Omega(k)}{k^2}, \quad p = 0, 1, \dots,$$

hence, putting $d := m + 2$ in the definition of δ_ν, we find the first estimate we want:

$$\frac{1}{4d}\sum_{\nu=0}^{\infty}\delta_\nu \le \sum_{\nu=0}^{\infty}\frac{\operatorname{Log}\Omega(2^{q+\nu+1})}{2^{q+\nu+2}} \le \sum_{\nu=0}^{\infty}\sum_{k=2^{q+\nu+1}}^{2^{q+\nu+2}-1}\frac{\operatorname{Log}\Omega(k)}{k^2} = \sum_{k=2^{q+1}}^{\infty}\frac{\operatorname{Log}\Omega(k)}{k^2}.$$

The second estimate follows from (15) and from the definitions of δ_ν, η_ν together with $t_\nu := s_{\nu+1} - 1$ and $g := m + 1$:

$$\begin{aligned}\eta_\nu &= \frac{1}{[\Omega(t_\nu)\delta_\nu]^g} \le \frac{1}{[t_\nu\delta_\nu]^g} \le \frac{1}{[s_\nu\delta_\nu]^g} = \frac{1}{[(m+2)\operatorname{Log}\Omega(t_\nu)]^g} \le \frac{1}{[3\operatorname{Log}t_\nu]^g}\\ &= \frac{1}{[(q+\nu+1)3\operatorname{Log}2]^g} \le \frac{1}{(q+\nu+1)^2} \le \frac{1}{(q+\nu)} - \frac{1}{(q+\nu+1)}.\end{aligned}$$

□

7. *Proof of Theorem 1.1*

First of all we construct a not necessarily area-preserving convergent transformation from (1) into (4) by means of Theorem 2.1. For this purpose we remark that the linear part of (1) is diagonalizable and has the eigenvalues

$$\lambda_1 = \lambda = e^{i\gamma_0} \ne 0, \quad \lambda_2 = \lambda^{-1} = e^{-i\gamma_0} \ne 0.$$

Moreover, as a consequence of (5)–(7), we have for $\lambda_1{}^{j_1}\lambda_2{}^{j_2} - \lambda_l \ne 0$ the relations

$$|\lambda_1{}^{j_1}\lambda_2{}^{j_2} - \lambda_l| = |\lambda^{|j_1-j_2\pm1|} - 1| \ge \frac{1}{\Omega(|j_1 - j_2 \pm 1|)} \ge \frac{1}{\Omega(j_1 + j_2 + 1)}$$

such that (14) and (15) are satisfied with $k \mapsto \Omega(k+1)$ instead of Ω. Therefore, Theorem 2.1 can be applied with $\mathbb{K} = \mathbb{R}$ and we get the desired real convergent change of variables (2) transforming (1) into (4), not necessarily being area preserving.

LEMMA 7.1. *If a transformation (2) from (1) into (4) is given without the condition* $\varphi_\xi\psi_\eta - \varphi_\eta\psi_\xi = 1$, *then we have*

$$(\varphi_\xi\psi_\eta - \varphi_\eta\psi_\xi)(\xi,\eta) = \sum_{k=0}^{\infty}\delta_k\left(\frac{\xi^2+\eta^2}{2}\right)^k, \quad \delta_0 = 1,\ \delta_k \in \mathbb{R},\ k = 1, 2, \dots . \tag{85}$$

Clearly, this series converges provided that φ *and* ψ *converge.*

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

Proof. We put

$$z = \begin{pmatrix} x \\ y \end{pmatrix}, \quad z_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \quad \zeta = \begin{pmatrix} \xi \\ \eta \end{pmatrix}, \quad \zeta_1 = \begin{pmatrix} \xi_1 \\ \eta_1 \end{pmatrix}, \quad S = \begin{pmatrix} \cos & -\sin \\ \sin & \cos \end{pmatrix} (\gamma_0) \tag{86}$$

and write $z_1 = h(z)$ for the mapping (1), $\zeta_1 = S\zeta$ for the mapping (4), and $z = \chi(\zeta)$, $z_1 = \chi(\zeta_1)$ for the change of variables (2). Then we obtain the equation $h(\chi(\zeta)) = \chi(S\zeta)$ from which we deduce the equation for the functional matrices h' of h and χ' of χ: $h'(\chi(\zeta))\chi'(\zeta) = \chi'(S\zeta)S$. Observing that h and $\zeta \mapsto S\zeta$ are area preserving and calculating the determinants of the members of the last equation we obtain

$$\Delta(\zeta) := \det \chi'(\zeta) = (\varphi_\xi \psi_\eta - \varphi_\eta \psi_\xi)(\xi, \eta) = \Delta(S\zeta). \tag{87}$$

With the matrix

$$T := \frac{1}{\sqrt{2}} \begin{pmatrix} i & 1 \\ 1 & i \end{pmatrix}$$

we get

$$\Lambda := \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} = T^* S T, \quad \lambda = e^{i\gamma_0}$$

and with the definition

$$D(\zeta) := \Delta(T\zeta) = \sum_{k,l \geq 0} D_{kl} \xi^k \eta^l \tag{88}$$

we obtain, using (87), $D(\Lambda\zeta) = \Delta(T\Lambda\zeta) = \Delta(ST\zeta) = \Delta(T\zeta) = D(\zeta)$. Hence,

$$0 = D(\Lambda\zeta) - D(\zeta) = \sum_{k,l \geq 0} (\lambda^{k-l} - 1) D_{kl} \xi^k \eta^l.$$

Since $\lambda^k \neq 1$ for $k \in \mathbb{Z}\backslash\{0\}$ by (5), the coefficients D_{kl} vanish for $k \neq l$. Therefore, we obtain $D(\zeta) = \sum_{k=0}^{\infty} D_{kk} (\xi\eta)^k$. This result leads, with (87) and (88), to the equations

$$(\varphi_\xi \psi_\eta - \varphi_\eta \psi_\xi)(\xi, \eta) = \Delta(\zeta) = D(T^*\zeta) = \sum_{k=0}^{\infty} D_{kk} (-i)^k \left(\frac{\xi^2 + \eta^2}{2} \right)^k.$$

$\varphi(\xi, \eta) = \xi + \cdots$ and $\psi(\xi, \eta) = \eta + \cdots$ are power series with real coefficients. Thus, (85) follows and the lemma is proved. □

LEMMA 7.2. *Let $M(\omega) := 1 + M_1\omega + M_2\omega^2 + \cdots$ be a formal power series in $\omega := \frac{1}{2}(\xi^2 + \eta^2)$ with real coefficients $M_1, M_2, \ldots$. Using the notation in (86), we form the change of variables*

$$z = \tilde{\chi}(\zeta) = \begin{pmatrix} \tilde{\varphi}(\xi, \eta) \\ \tilde{\psi}(\xi, \eta) \end{pmatrix} = M(\omega)\zeta, \quad z_1 = M(\omega_1)\zeta_1, \quad \omega_1 := \frac{1}{2}(\xi_1^2 + \eta_1^2). \tag{89}$$

Then the rotation $z \mapsto z_1 = Sz$ is transformed by means of this change of variables into itself: $\zeta_1 = S\zeta$. The angle γ_0 in S may be an arbitrary real number here. In addition, we have

$$\det \tilde{\chi}'(\zeta) = (\tilde{\varphi}_\xi \tilde{\psi}_\eta - \tilde{\varphi}_\eta \tilde{\psi}_\xi)(\xi, \eta) = \frac{d}{d\omega}[\omega M(\omega)^2]. \tag{90}$$

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

Proof. Replacing z, z_1 in $z_1 = Sz$ by ζ_1, ζ, using (89), we obtain

$$M(\omega_1)\zeta_1 = SM(\omega)\zeta. \tag{91}$$

Multiplication of this equation from the left with its transpose yields $M(\omega_1)^2\omega_1 = M(\omega)^2\omega = \omega + \cdots$ or $0 = M(\omega_1)^2\omega_1 - M(\omega)^2\omega = (\omega_1 - \omega)(1 + \cdots)$; hence, $\omega_1 = \omega$. Therefore, from (91) it follows that $\zeta_1 = S\zeta$. From (89) we see that $\tilde{\varphi} = M(\omega)\xi$, $\tilde{\psi} = M(\omega)\eta$ such that a simple calculation yields (90): $(\tilde{\varphi}_\xi \tilde{\psi}_\eta - \tilde{\varphi}_\eta \tilde{\psi}_\xi) = (M + \xi^2 M_\omega)(M + \eta^2 M_\omega) - \xi^2\eta^2(M_\omega)^2 = M^2 + MM_\omega(\xi^2 + \eta^2) = (\omega M^2)_\omega$. □

With the two lemmata above we are now in a position to construct a convergent area-preserving change of variables (2) transforming (1) into (4). First of all we take the convergent change of variables

$$z = \chi(\zeta) = \begin{pmatrix} \varphi \\ \psi \end{pmatrix}(\xi, \eta)$$

transforming (1) into (4), which is guaranteed by Theorem 2.1, as mentioned above, and to which Lemma 7.1 can be applied. This transformation has to be combined with a change of variables $z = \tilde{\chi}(\zeta) = M(\omega)\zeta$ leaving invariant rotations like (4), as stated in Lemma 7.2. So we form

$$z = \chi \circ \tilde{\chi}(\zeta) = \chi(M(\omega)\zeta) \tag{92}$$

trying to determine $M(\omega) = 1 + \cdots$ as a convergent power series of $\omega = \frac{1}{2}(\xi^2 + \eta^2)$ in such a way that this change of variables transforming (1) into (4) is area preserving. With the functional matrix χ' of χ and $\tilde{\chi}'$ of $\tilde{\chi}$, we first obtain from (85) $\det \chi'(\zeta) = F(\omega) := 1 + \delta_1\omega + \delta_2\omega^2 + \cdots$. Hence,

$$\det \chi'(\tilde{\chi}(\zeta)) = F(\omega M(\omega)^2) \tag{93}$$

where the power series F is convergent. Then we get, using (90), (92) and (93),

$$\det(\chi \circ \tilde{\chi})'(\zeta) = \det \chi'(\tilde{\chi}(\zeta)) \det \tilde{\chi}'(\zeta) = F(\omega M(\omega)^2)\frac{d}{d\omega}[\omega M(\omega)^2] = 1 \tag{94}$$

where the last equation has to be satisfied by means of a proper convergent power series $M(\omega) = 1 + \cdots$ in ω.

Putting $v = \omega M(\omega)^2$, equation (94) reads $d\omega = F(v)\,dv$ with the convergent solution

$$\omega = \int_0^v F(u)\,du = v + \tfrac{1}{2}\delta_1 v^2 + \tfrac{1}{3}\delta_2 v^3 + \cdots$$

having a convergent inverse $v = H(\omega) := \omega - \frac{1}{2}\delta_1\omega^2 + (\frac{1}{2}\delta_1^2 - \frac{1}{3}\delta_2)\omega^3 + \cdots$. Now M has to be determined from the equation $\omega M(\omega)^2 = H(\omega)$ which has the convergent solution

$$M(\omega) = \left(\frac{H(\omega)}{\omega}\right)^{1/2} = (1 + \cdots)^{1/2} = 1 + \cdots .$$

Thus, a convergent area-preserving change of variables (2) from (1) into (4) has been found. Theorem 1.1 is proved. □

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

8. *The Bruno condition*

Preliminaries. For the irrational number $\gamma_0/2\pi$ in §1 we define

$$\alpha := \min_{l\in\mathbb{Z}} \left|\frac{\gamma_0}{2\pi} - l\right|$$

such that we get

$$0 < \alpha < \tfrac{1}{2}, \quad \lambda = e^{i\gamma_0} = e^{\pm 2\pi i\alpha}. \tag{95}$$

α is also irrational. From the elementary fact $(2/\pi)x < \sin x < x$ for $0 < x < (\pi/2)$, we deduce for

$$0 < D_k := \min_{l\in\mathbb{Z}} |k\alpha - l| < \tfrac{1}{2}, \quad k = 1, 2, \ldots \tag{96}$$

the inequalities

$$4D_k < 2\sin\pi D_k = |\lambda^k - 1| < \pi D_k, \quad k = 1, 2, \ldots .$$

This implies the existence of a function $\Omega : \mathbb{N} \to \mathbb{R}$ satisfying (5), (6) and (7), is equivalent to the existence of such a function satisfying

$$D_k \geq \frac{1}{\Omega(k)}, \quad k = 1, 2, \ldots, \tag{97}$$

and equations (6) and (7). For the formulation of the Bruno condition [**1**] and the proof that it is equivalent to the existence of such a function Ω we need the main denominators $1 = q_0 < q_1 < \cdots$ of the continued fraction of α and the following of its properties:

$$D_k \geq D_{q_n} > D_{q_{n+1}} \quad \text{for } 1 \leq k < q_{n+1},\ n = 0, 1, \ldots, \tag{98}$$

$$q_{n+1} < \frac{1}{D_{q_n}} < 2q_{n+1} < q_{n+1}^2, \quad q_n \geq 2^{n/2} \quad \text{for } n = 0, 1, \ldots . \tag{99}$$

Definition 8.1. A number $\alpha \in]0, \frac{1}{2}[\setminus \mathbb{Q}$ satisfies the Bruno condition if the estimate

$$\sum_{n=0}^{\infty} \frac{\operatorname{Log} q_{n+1}}{q_n} < \infty \tag{100}$$

holds for the main denominators $1 = q_0 < q_1 < \cdots$ of the continued fraction of α.

LEMMA 8.1. *$\Omega^* : \mathbb{N} \to \mathbb{R}$ defined by $\Omega^*(k) := (1/D_{q_n})$, $q_n \leq k < q_{n+1}$, $n = 0, 1, \ldots$ has the properties*

$$D_k \geq \frac{1}{\Omega^*(k)}, \quad k = 1, 2, \ldots, \tag{101}$$

$$k < \Omega^*(k) \leq \Omega^*(k+1), \quad k = 1, 2, \ldots . \tag{102}$$

Proof. Equation (98) leads to $D_k \geq D_{q_n} = 1/\Omega^*(k)$ for $q_n \leq k < q_{n+1}$, $n = 0, 1, \ldots$ and, hence, to (101). The second inequality in (102) is also a consequence of (98). Finally, we get $\Omega^*(k) = {D_{q_n}}^{-1} > q_{n+1} > k \geq q_n$, $n = 0, 1, \ldots$ by (99). The lemma is proved. □

LEMMA 8.2.

$$\sum_{n=0}^{\infty} \frac{\operatorname{Log} q_{n+1}}{q_n} < \infty \quad \Longleftrightarrow \quad \sum_{k=1}^{\infty} \frac{\operatorname{Log} \Omega^*(k)}{k^2} < \infty.$$

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

Proof. With the definition of Ω^* we find:
(1)

$$\begin{aligned}\sum_{k=q_1}^{\infty}\frac{\operatorname{Log}\Omega^*(k)}{k^2} &= \sum_{n=1}^{\infty}\operatorname{Log}\frac{1}{D_{q_n}}\sum_{k=q_n}^{q_{n+1}-1}\frac{1}{k^2}\\ &\le \sum_{n=1}^{\infty}\operatorname{Log}\frac{1}{D_{q_n}}\sum_{k=q_n}^{q_{n+1}-1}\left(\frac{1}{k-1}-\frac{1}{k}\right)\\ &= \sum_{n=1}^{\infty}\operatorname{Log}\frac{1}{D_{q_n}}\left(\frac{1}{q_n-1}-\frac{1}{q_{n+1}-1}\right)\\ &\le \sum_{n=1}^{\infty}\operatorname{Log} q_{n+1}\frac{4}{q_n}\end{aligned}$$

because of $1/D_{q_n} < q_{n+1}^2$ in (99) and

$$\frac{1}{q_n-1}\le\frac{2}{q_n},\quad n\ge 1;$$

(2)

$$\begin{aligned}\sum_{k=q_1}^{\infty}\frac{\operatorname{Log}\Omega^*(k)}{k^2} &= \sum_{n=1}^{\infty}\operatorname{Log}\frac{1}{D_{q_n}}\sum_{k=q_n}^{q_{n+1}-1}\frac{1}{k^2}\\ &\ge \sum_{n=1}^{\infty}\operatorname{Log}\frac{1}{D_{q_n}}\sum_{k=q_n}^{q_{n+1}-1}\left(\frac{1}{k}-\frac{1}{k+1}\right)\\ &= \sum_{n=1}^{\infty}\operatorname{Log}\frac{1}{D_{q_n}}\left(\frac{1}{q_n}-\frac{1}{q_{n+1}}\right)\\ &\ge \sum_{n=1}^{\infty}\operatorname{Log} q_{n+1}\left(\frac{1}{q_n}-\frac{1}{q_{n+1}}\right)\end{aligned}$$

because of $1/D_{q_n} > q_{n+1}$ in (99).

The chain of inequalities (1) proves the implication $\Longrightarrow$. Chain (2) is still not complete. Since $x\mapsto x^{-1}\operatorname{Log} x$ is decreasing for $x\ge e$, we have $q_n^{-1}\operatorname{Log} q_n \le 2^{-n/2}\operatorname{Log} 2^{n/2}$ for $n\ge 3$, in view of (99). Therefore, the series $\sum_{n=2}^{\infty} {q_n}^{-1}\operatorname{Log} q_n$ converges such that we may add it to chain of inequalities (2) above in order to obtain the inequality

$$\sum_{n=1}^{\infty}\frac{\operatorname{Log} q_{n+1}}{q_n}\le\sum_{k=q_1}^{\infty}\frac{\operatorname{Log}\Omega^*(k)}{k^2}+\sum_{n=2}^{\infty}\frac{\operatorname{Log} q_n}{q_n}$$

from which the implication $\Longleftarrow$ follows. The lemma is proved. □

LEMMA 8.3. *Any function $\Omega:\mathbb{N}\to\mathbb{R}$ satisfying the inequalities (101) and (102) with Ω instead of Ω^* fulfills the inequalities*

$$\Omega(k)\ge\Omega^*(k),\quad k=1,2,\ldots. \tag{103}$$

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press

Proof. Otherwise there is some $k^* \in \mathbb{N}$ with $\Omega(k^*) < \Omega^*(k^*)$ and then there is some $n^* \in \mathbb{N}$ with $q_{n^*} \leq k^* < q_{n^*+1}$. As a result we obtain the contradiction

$$D_{q_{n^*}} = \frac{1}{\Omega^*(k^*)} < \frac{1}{\Omega(k^*)} \leq \frac{1}{\Omega(q_{n^*})} \leq D_{q_{n^*}},$$

using the definition of Ω^* and the properties of Ω. The lemma is proved. □

Lemma 8.1 and Lemma 8.2 show that the Bruno condition is equivalent to the function Ω^* satisfying (101), (102) and (7), with $\Omega = \Omega^*$. It follows, with Lemma 8.3, that any function Ω satisfying (97), (6) and (7) causes the validity of the Bruno condition. Together with the remarks in the Preliminaries we have shown that the existence of a function Ω satisfying (5)–(7) is equivalent to the Bruno condition.

References

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

[2] C. Camacho, N. H. Kuiper and J. Palis. The topology of holomorphic flows with singularity. *Inst. Hautes Etudes Sci. Publ. Math.* **48** (1978), 5–38.

[3] M. Herman. Recent results and some open questions on Siegel's linearization theorem of germs of complex analytic diffeomorphisms of $\mathbb{C}^n$ near a fixed point. *Proc. VIII Int. Conf. on Mathematics and Physics (Marseille 1986)*. World Scientific Publishing, Singapore, 1987, pp. 138–198.

[4] M. Herman. Some open problems in dynamical systems. *Proc. ICM Berlin*, Vol. II, 1998, pp. 797–808.

[5] H. Rüssmann. Über die Normalform analytischer Hamiltonscher Differentialgleichungen in der Nähe einer Gleichgewichtslösung. *Math. Ann.* **169** (1967), 55–72.

[6] H. Rüssmann. On the convergence of power series transformations of analytic mappings near a fixed point into a normal form. *Preprint I.H.E.S.* **M/77/178** (1977), 1–44.

[7] C. L. Siegel and J. K. Moser. *Lectures on Celestial Mechanics*. Springer, Berlin, 1971.

[8] E. Zehnder. *A Simple Proof of a Generalization of a Theorem by C. L. Siegel* (*Lecture Notes in Mathematics, 597*). Springer, Berlin, 1977, pp. 828–854.

https://doi.org/10.1017/S0143385702000974 Published online by Cambridge University Press