

Note on linearization

by MAREK KUCZMA (Katowice)

Abstract. Under suitable conditions the existence and uniqueness of smooth solutions g is proved for the functional equation $g[f(x)] = Sg(x)$. Here f, g, x belong to \mathbf{R}^n and S is a real $n \times n$ matrix.

1. When one considers an autonomous system of differential equations

$$(1) \quad \dot{x} = a(x),$$

the linearization turns out to be a useful device allowing one to reduce the study of the solutions of (1) to the study of the solutions of the corresponding linear system. The procedure is described in detail in [8] and in [3], chapter IX. An essential step is to know that the functional equation ⁽¹⁾

$$(2) \quad g[f(x)] = Sg(x)$$

(where the properties of the function f are determined by those of a in (1)) has a locally invertible solution g ; and it is desirable to have g as smooth as possible.

In the present note we are going to prove a theorem to this effect. The theorem improves slightly on an analogous result of S. Sternberg [8].

2. We shall adopt the following notational conventions.

$n \geq 1$ is a fixed integer. Lower case Latin letters from i to r denote positive integers; letters up to h and from s on denote elements of \mathbf{R}^n . Capital Latin letters denote square $n \times n$ matrices with real entries. Lower case Greek letters denote real numbers, except for σ which may be real or complex. Differentiation is marked by lower indices; thus $g_{i_1 \dots i_q}(x)$ denotes

$$\frac{\partial^q g}{\partial \xi^{i_1} \dots \partial \xi^{i_q}}(x), \quad x = (\xi^1, \dots, \xi^n).$$

⁽¹⁾ Equation (2) may be regarded as a generalization of the Schröder equation; cf. [5], [6].

In particular, for the function $f = (\varphi^1, \dots, \varphi^n)$ we write

$$\frac{\partial^a \varphi^k}{\partial \xi^{i_1} \dots \partial \xi^{i_q}}(x) = \varphi_{i_1 \dots i_q}^k(x).$$

S will always denote the matrix

$$(3) \quad S = (\varphi_i^k(0)),$$

and $\sigma_1, \dots, \sigma_n$ denote the eigenvalues of S . These eigenvalues will be subjected to the condition

$$(4) \quad 0 < |\sigma_1| \leq \dots \leq |\sigma_n| < 1.$$

For $x \in \mathbf{R}^n$, $|x|$ denotes the usual Euclidean norm of x .

We introduce also a sequence of spaces $\mathbf{X}^{(q)}$, $q = 1, 2, \dots$. The elements of $\mathbf{X}^{(q)}$ (which will be denoted by bold face, lower case Latin letters) are the n^q -tuples of elements of \mathbf{R}^n :

$$\mathbf{u}^{(q)} = \{u_{i_1 \dots i_q}\}, \quad i_1, \dots, i_q = 1, \dots, n.$$

The linear structure in $\mathbf{X}^{(q)}$ is introduced in the obvious way; the norm is defined by

$$\|\mathbf{u}^{(q)}\|_q = \left(\sum_{i_1, \dots, i_q=1}^n u_{i_1 \dots i_q}^2 \right)^{1/2}.$$

The norm of a matrix $B = (\beta_i^j)$ is defined as

$$\|B\| = \sup_{\|\mathbf{u}^{(1)}\|_1=1} \|B^T \mathbf{u}^{(1)}\|_1,$$

B^T being the transpose of B so that $B^T \mathbf{u}^{(1)} = \left\{ \sum_{j=1}^n \beta_i^j u_j \right\}$. This norm has in particular the property ([5], Lemma 6.2)

$$(5) \quad |Bx| \leq \|B\| |x| \quad \text{for } x \in \mathbf{R}^n.$$

It follows from (5) that we have for $\mathbf{u}^{(q)} = \{u_{i_1 \dots i_q}\} \in \mathbf{X}^{(q)}$

$$(6) \quad \|\{B u_{i_1 \dots i_q}\}\|_q \leq \|B\| \|\{u_{i_1 \dots i_q}\}\|_q.$$

3. The following theorem ([8], [2], [3] chapter IX; a similar result can be deduced also from [5]) deals with the existence of locally invertible smooth solutions of equation (2) in the case where f is a local contraction.

THEOREM 1. *Suppose that f is defined and of class C^r , $r \geq 1$, in a neighbourhood of the origin, with*

$$(7) \quad f(0) = 0,$$

and let the eigenvalues of matrix (3) fulfil condition (4) and

$$(8) \quad \sigma_1^{m_1} \dots \sigma_n^{m_n} \neq \sigma_j, \quad j = 1, \dots, n,$$

for all systems of non-negative integers m_1, \dots, m_n satisfying $1 < m_1 + \dots + m_n \leq r$. If

$$|\sigma_n|^r / |\sigma_1| < 1,$$

then equation (2) has a solution $g = (\gamma^1, \dots, \gamma^n)$ which is defined and of class C^r in a neighbourhood of the origin and such that the matrix $(\gamma_i^j(0))$ is non-singular.

The purpose of the present note is to prove a modified version of the above theorem. The method of proof will follow the lines of [5] and will be different from that employed in [8], [2], [3].

First let us note the following lemma ([5], Lemma 2.1).

LEMMA 1. If f and g are of class C^r in a region Δ such that $f(\Delta) \subset \Delta$, then we have for $x \in \Delta$

$$\frac{\partial^q}{\partial \xi^{i_1} \dots \partial \xi^{i_q}} g([f(x)]) = \sum_{k=1}^q \sum_{j_1, \dots, j_k=1}^n \beta_{i_1 \dots i_q}^{j_1 \dots j_k}(x) g_{j_1 \dots j_k}[f(x)], \quad q = 1, \dots, r,$$

where the functions $\beta_{i_1 \dots i_q}^{j_1 \dots j_k}$ may be expressed by means of sums and products of $\varphi_i^j, \dots, \varphi_{i_1 \dots i_{q-k+1}}^j$. Consequently $\beta_{i_1 \dots i_q}^{j_1 \dots j_k}$ are of class $C^{r-q+k-1}$ in Δ . In particular,

$$\beta_{i_1 \dots i_q}^{j_1 \dots j_q}(x) = \varphi_{i_1}^{j_1}(x) \dots \varphi_{i_q}^{j_q}(x).$$

Differentiating equation (2) and then setting $x = 0$ (observe (7)) we see that the values

$$(9) \quad v_{i_1 \dots i_q} = g_{i_1 \dots i_q}(0), \quad i_1, \dots, i_q = 1, \dots, n; \quad q = 1, \dots, r,$$

must fulfil the system of linear equations

$$(10) \quad S v_{i_1 \dots i_q} = \sum_{k=1}^q \sum_{j_1, \dots, j_k=1}^n \beta_{i_1 \dots i_q}^{j_1 \dots j_k}(0) v_{j_1 \dots j_k},$$

$$i_1, \dots, i_q = 1, \dots, n; \quad q = 1, \dots, r.$$

Moreover, $v_{i_1 \dots i_q}$ must fulfil also the consistency conditions

$$(11) \quad v_{i_{l_1} \dots i_{l_q}} = v_{i_1 \dots i_q}$$

for every permutation l_1, \dots, l_q of the sequence $1, \dots, q$. Systems

$$(12) \quad v_{i_1 \dots i_q}, \quad i_1, \dots, i_q = 1, \dots, n; \quad q = 1, \dots, r$$

fulfilling equations (10) and consistency conditions (11) will be called *admissible*.

Condition (8) guarantees the existence of admissible systems such that v_i form a non-singular matrix (cf. [7]). But there may exist admissible systems with this property also if condition (8) is not fulfilled.

We aim at proving the following

THEOREM 2. *Suppose that f is defined and of class C^r , $r \geq 1$, in a neighbourhood of the origin, with (7), and let the eigenvalues of matrix (3) fulfil condition (4). Further let*

$$f_{i_1 \dots i_r}(x) = f_{i_1 \dots i_r}(0) + O(|x|^\delta), \quad x \rightarrow 0; \quad 0 \leq \delta \leq 1,$$

$$i_1, \dots, i_r = 1, \dots, n.$$

If

$$(13) \quad |\sigma_n|^{r+\delta}/|\sigma_1| < 1,$$

then for every system of admissible values (12) equation (2) has a unique solution g which is defined and of class C^r in a neighbourhood of the origin, fulfils (9), and

$$(14) \quad g_{i_1 \dots i_r}(x) = g_{i_1 \dots i_r}(0) + O(|x|^\delta), \quad x \rightarrow 0,$$

$$i_1, \dots, i_r = 1, \dots, n.$$

We obtain Theorem 1 as a particular case $\delta = 0$ of Theorem 2.

4. In the proof of Theorem 2 we shall make use of the following lemma ([5], Lemma 6.3).

LEMMA 2. *If $B = (\beta_i^j)$, $\mathbf{s}^{(a)} = \{s_{i_1 \dots i_q}\}$, $\mathbf{t}^{(a)} = \{t_{i_1 \dots i_q}\}$, $\|B\| = \lambda$ and*

$$t_{i_1 \dots i_q} = \sum_{j_1, \dots, j_q=1}^n \beta_{i_1}^{j_1} \dots \beta_{i_q}^{j_q} s_{j_1 \dots j_q},$$

then

$$\|\mathbf{t}^{(a)}\|_q \leq \lambda^q \|\mathbf{s}^{(a)}\|_q.$$

Next note that, given an $\varepsilon > 0$, we may find a non-singular matrix T such that

$$\|T^{-1}ST\| \leq |\sigma_n| + \varepsilon \quad \text{and} \quad \|T^{-1}S^{-1}T\| \leq |\sigma_1|^{-1} + \varepsilon$$

(this may be achieved in a similar way to that described in [4], § 4). Now, if we put $f^*(x) = T^{-1}f(Tx)$ and $g^*(x) = T^{-1}g(Tx)$, then equation (2) goes into

$$g^*[f^*(x)] = T^{-1}STg^*(x).$$

Therefore we may assume that the norms

$$\lambda = \|S\| \quad \text{and} \quad \mu = \|S^{-1}\|$$

fulfil the condition

$$(15) \quad \lambda^{r+\delta} \mu < 1.$$

It follows from (15) that $\lambda < 1$. For every a fulfilling $\lambda < a < 1$ we have

$$(16) \quad |f(x)| \leq a|x|$$

and

$$(17) \quad \|\varphi_i^j(x)\| \leq a$$

for $|x|$ sufficiently small.

We choose an a , $\lambda < a < 1$, such that

$$(18) \quad a^{r+\delta} \mu < 1.$$

Let $\eta > 0$ be such that the function f is defined and of class C^r in $\Delta_\eta = \{x: |x| \leq \eta\}$ and that relations (16) and (17) hold in Δ_η . (Later η will be subjected to further restrictions.) For a fixed system of admissible values (12) we define \mathcal{D} as the space of all functions g which are defined and of class C^r in Δ_η , vanish at the origin, and satisfy conditions (9) and (14).

In \mathcal{D} we introduce the metric

$$(19) \quad \varrho(g, h) = \sup_{0 < |x| \leq \eta} |x|^{-\delta} \|\mathbf{g}^{(r)}(x) - \mathbf{h}^{(r)}(x)\|_r,$$

where we have adopted the notation

$$\mathbf{g}^{(q)}(x) = \{g_{i_1 \dots i_q}(x)\}, \quad \mathbf{h}^{(q)}(x) = \{h_{i_1 \dots i_q}(x)\}, \quad q = 1, \dots, r.$$

We have for $0 < |x| \leq \eta$ and $q = 1, \dots, r-1$

$$\|\mathbf{g}^{(q)}(x) - \mathbf{h}^{(q)}(x)\|_q \leq \sup_{0 < |y| \leq |x|} \|\mathbf{g}^{(q+1)}(y) - \mathbf{h}^{(q+1)}(y)\|_{q+1} |x|.$$

But for $0 < |y| \leq |x|$

$$\|\mathbf{g}^{(q+1)}(y) - \mathbf{h}^{(q+1)}(y)\|_{q+1} \leq |y|^{-\delta} \|\mathbf{g}^{(q+1)}(y) - \mathbf{h}^{(q+1)}(y)\|_{q+1} |x|^\delta,$$

whence

$$\sup_{0 < |x| \leq \eta} |x|^{-\delta} \|\mathbf{g}^{(q)}(x) - \mathbf{h}^{(q)}(x)\|_q \leq \eta \sup_{0 < |y| \leq \eta} |y|^{-\delta} \|\mathbf{g}^{(q+1)}(y) - \mathbf{h}^{(q+1)}(y)\|_{q+1}.$$

Hence we get

$$(20) \quad \sup_{0 < |x| \leq \eta} |x|^{-\delta} \|\mathbf{g}^{(q)}(x) - \mathbf{h}^{(q)}(x)\|_q \leq \eta^{r-q} \varrho(g, h), \quad q = 1, \dots, r-1,$$

and similarly

$$(21) \quad \sup_{0 < |x| \leq \eta} |x|^{-\delta} |g(x) - h(x)| \leq \eta^r \varrho(g, h).$$

Relations (20) and (21) show that the convergence in the space (\mathcal{D}, ϱ) is equivalent to the uniform convergence in Δ_η of the sequence of functions together with their derivatives up to the order r . Consequently \mathcal{D} endowed with metric (19) is a complete metric space.

Now consider the transform $\mathfrak{I}: \mathscr{D} \rightarrow \mathscr{D}$ defined by

$$\mathfrak{I}(g)(x) = S^{-1}g[f(x)].$$

It is a straightforward matter to check that \mathfrak{I} actually maps \mathscr{D} into itself, and we will be through when we show that \mathfrak{I} is a contraction map.

Take arbitrary $g, h \in \mathscr{D}$ and put $\tilde{g} = \mathfrak{I}(g)$, $\tilde{h} = \mathfrak{I}(h)$. We have by Lemma 1

$$\begin{aligned} \tilde{g}_{i_1 \dots i_r}(x) - \tilde{h}_{i_1 \dots i_r}(x) \\ = S^{-1} \sum_{k=1}^r \sum_{j_1, \dots, j_k=1}^n \beta_{i_1 \dots i_r}^{j_1 \dots j_k}(x) (g_{j_1 \dots j_k}[f(x)] - h_{j_1 \dots j_k}[f(x)]), \end{aligned}$$

whence by (6)

$$\begin{aligned} \|\tilde{g}^{(r)}(x) - \tilde{h}^{(r)}(x)\|_r \\ \leq \mu \sum_{k=1}^r \left\| \left\{ \sum_{j_1, \dots, j_k=1}^n \beta_{i_1 \dots i_r}^{j_1 \dots j_k}(x) (g_{j_1 \dots j_k}[f(x)] - h_{j_1 \dots j_k}[f(x)]) \right\} \right\|_r. \end{aligned}$$

Since the functions β are at least continuous in Δ_η , we have for $k < r$

$$\begin{aligned} \sup_{0 < |x| \leq \eta} |x|^{-\delta} \left\| \left\{ \sum_{j_1, \dots, j_k=1}^n \beta_{i_1 \dots i_r}^{j_1 \dots j_k}(x) (g_{j_1 \dots j_k}[f(x)] - h_{j_1 \dots j_k}[f(x)]) \right\} \right\|_r \\ \leq \omega_k \sup_{0 < |x| \leq \eta} |x|^{-\delta} \|g^{(k)}[f(x)] - h^{(k)}[f(x)]\|_k, \end{aligned}$$

where the coefficients ω_k depend only on β (and thus on f), but not on the particular functions g, h . Further we get by (20) and (16)

$$\begin{aligned} \sup_{0 < |x| \leq \eta} |x|^{-\delta} \|g^{(k)}[f(x)] - h^{(k)}[f(x)]\|_k \\ \leq \sup_{0 < |x| \leq \eta} (|f(x)|/|x|)^\delta |f(x)|^{-\delta} \|g^{(k)}[f(x)] - h^{(k)}[f(x)]\|_k \\ \leq \alpha^\delta \sup_{0 < |x| \leq \eta} |x|^{-\delta} \|g^{(k)}(x) - h^{(k)}(x)\|_k \leq \alpha^\delta \eta^{r-k} \varrho(g, h). \end{aligned}$$

For $k = r$ we have in view of Lemmas 1 and 2 and of relations (16) and (17)

$$\begin{aligned} \sup_{0 < |x| \leq \eta} |x|^{-\delta} \left\| \left\{ \sum_{j_1, \dots, j_r=1}^n \beta_{i_1 \dots i_r}^{j_1 \dots j_r}(x) (g_{j_1 \dots j_r}[f(x)] - h_{j_1 \dots j_r}[f(x)]) \right\} \right\|_r \\ \leq \alpha^\delta \alpha^r \varrho(g, h). \end{aligned}$$

Finally we get the estimation

$$\begin{aligned} \varrho(\tilde{g}, \tilde{h}) &= \sup_{0 < |x| \leq \eta} |x|^{-\delta} \|\tilde{g}^{(r)}(x) - \tilde{h}^{(r)}(x)\|_r \\ &\leq (\alpha^{r+\delta} \mu + \alpha^\delta \mu \sum_{k=1}^{r-1} \omega_k \eta^{r-k}) \varrho(g, h). \end{aligned}$$

In view of (18) we can make the expression in the parentheses less than 1 choosing η sufficiently small. Thus \mathfrak{I} is a contraction provided η has been chosen sufficiently small, which was to be proved.

The solution has been obtained locally, in a neighbourhood of the origin. It may be extended onto a larger domain (the domain of attraction of 0; cf. [6]) by a use of equation (2).

The solution may be obtained as the limit of a sequence of successive approximations.

5. In the one-dimensional case ($n = 1$) the result obtained is closely related to the following theorem due to Szekeres [9] and Crum [1] (cf. also [6], Theorem 6.3).

THEOREM 3. *If f is defined and of class C^1 in a neighbourhood of the origin and satisfies (7) and*

$$f'(x) = s + O(|x|^\delta), \quad x \rightarrow 0; \quad \delta > 0,$$

where $0 < s < 1$, then for every real v there exists exactly one solution g of the equation

$$g[f(x)] = sg(x)$$

which is defined and of class C^1 in a neighbourhood of the origin and fulfils $g'(0) = v$.

Now, from Theorem 2 above we receive an additional information that this solution fulfils the condition

$$g'(x) = v + O(|x|^\delta), \quad x \rightarrow 0.$$

Theorem 2 may be regarded as a joint generalization of Theorems 1 and 3 and yields a new proof of those theorems.

References

- [1] M. Crum, *On two functional equations which occur in the theory of clock-graduation*, Quart. J. Math. Oxford Ser. 10 (1939), p. 155-160.
- [2] Ph. Hartman, *On local homeomorphisms of Euclidean spaces*, Bol. Soc. Mat. Mexicana (2) 5 (1960), p. 220-241.
- [3] — *Ordinary differential equations*, New York — London — Sydney 1964.
- [4] J. Kordylewski, *On continuous solutions of systems of functional equations*, Ann. Polon. Math. 25 (1971), p. 53-83.
- [5] M. Kuczma, *On the Schröder equation*, Rozprawy Mat. 34 (1963).
- [6] — *Functional equations in a single variable*, Monografie Mat. vol. 46, Warszawa 1968.
- [7] W. Smajdor, *Formal solutions of a functional equation*, Zeszyty Naukowe Uniwersytetu Jagiellońskiego, Prace Mat. 13 (1969), p. 71-78.
- [8] S. Sternberg, *Local contractions and a theorem of Poincaré*, Amer. J. Math. 79 (1957), p. 809-824.
- [9] G. Szekeres, *Regular iteration of real and complex functions*, Acta Math. 100 (1958), p. 203-258.

Reçu par la Rédaction le 14. 2. 1973

