

ANNALI DELLA  
SCUOLA NORMALE SUPERIORE DI PISA  
*Classe di Scienze*

JÜRGEN MOSER

**A rapidly convergent iteration method and non-linear  
differential equations = II**

*Annali della Scuola Normale Superiore di Pisa, Classe di Scienze 3<sup>e</sup> série*, tome 20,  
n° 3 (1966), p. 499-535

[http://www.numdam.org/item?id=ASNSP\\_1966\\_3\\_20\\_3\\_499\\_0](http://www.numdam.org/item?id=ASNSP_1966_3_20_3_499_0)

© Scuola Normale Superiore, Pisa, 1966, tous droits réservés.

L'accès aux archives de la revue « *Annali della Scuola Normale Superiore di Pisa, Classe di Scienze* » (<http://www.sns.it/it/edizioni/riviste/annaliscienze/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

*Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques*  
<http://www.numdam.org/>

# A RAPIDLY CONVERGENT ITERATION METHOD AND NON-LINEAR DIFFERENTIAL EQUATIONS = II

JÜRGEN MOSER (\*)

## CHAPTER. III — Conjugacy Problems.

The method explained above is, of course, not only applicable to partial differential equations but to many other functional equations

$$\mathcal{F}(u) = f.$$

The essential requirements for this method are *a)* that an approximate solution, say  $u = u_0$ , is known, so that  $\mathcal{F}(u_0)$  is sufficiently close to  $f$  and *b)* that the linearized equation

$$(1) \quad \mathcal{F}'(w)v = g$$

admits a solution for  $v$  — for any given  $w$  in a neighborhood of  $u_0$ . This second condition is rather stringent and often it is possible to ensure the solvability of the equation

$$(2) \quad \mathcal{F}'(u_0)v = g$$

but not that of (1). In fact, in this chapter we shall discuss just such problems for which (1) is not solvable but (2) is.

Such problems occur in celestial mechanics and are closely related to the so-called small divisor difficulty. We shall discuss the simplest model problem in which this difficulty occurs, the center problem. This problem was treated successfully by C. L. Siegel in [5]. For a detailed discussion

---

Pervenuto alla Redazione il 23 Giugno 1965.

(\*) The first part has appeared on the January/March 1966 issue of this same Journal.

of this problem we refer to the book [22], Section 23 and Section 24 and for some similar results for differential equations we mention [5'].

We shall describe this problem in the following section but approach it with a different method which allows other applications. In fact, this is exactly the method which was suggested by Kolmogorov [6] for a small divisor problem and was used by Arnold [10] in the proof of Kolmogorov's theorem. The main point will be to construct a rapidly convergent sequence of approximations by solving the linearized equations (2) (linearized at  $w = u_0$ ) only. For general functional equations this is not possible but for the problems discussed here (conjugacy problem) we shall describe such a method (see Section 2).

### § 1. Siegel's theorem.

We consider a conformal mapping

$$z \rightarrow z_1 = f(z)$$

near a fixed point, say  $z = 0$ . Hence  $f(0) = 0$  and we can write

$$(1.1) \quad z_1 = \lambda z + \widehat{f}(z)$$

where  $\widehat{f}(z)$  is a power series which vanishes quadratically at  $z = 0$ . We assume  $\lambda \neq 0$ .

The problem to be discussed is to find a coordinate transformation

$$(1.2) \quad z = u(\zeta) = \zeta + \widehat{u}(\zeta)$$

— where  $\widehat{u}(\zeta)$  again vanishes quadratically at  $\zeta = 0$  — such that in the coordinates  $\zeta$  the mapping (1.1) takes the simple linear form

$$\zeta_1 = \lambda \zeta.$$

The mapping (1.1) is called «conjugate» to a linear one if such a transformation (1.2) can be found.

It is very easy to determine the coefficients of  $u$  by formal expansion provided  $\lambda$  is not a root of unity. Indeed the functional equation which has to be solved is

$$(1.3) \quad u(\lambda \zeta) = f(u(\zeta)).$$

Assuming that in

$$u(\zeta) = \zeta + u_2 \zeta^2 + \dots$$

the coefficients  $u_2, u_3, \dots, u_{k-1}$  have been found in such a way as to satisfy (1.3) (mod  $\zeta^k$ ) we find for the coefficient of  $\zeta^k$  from (1.3)

$$u_k \lambda^k \zeta^k - \lambda u_k \zeta^k = g_k \zeta^k, \quad (k = 2, 3, \dots).$$

Here  $g_k$  is known, as it depends on  $u_2, u_3, \dots, u_{k-1}$  and the coefficients of  $f$  only. Hence  $u_k$  is determined by the equation

$$(1.4) \quad (\lambda^k - \lambda) u_k = g_k$$

which is uniquely solvable if  $\lambda$  is not a unit root.

The question we are concerned with is the convergence of this series for  $u$ . If  $|\lambda| \neq 1$  the convergence can be established straightforwardly by Cauchy's majorant method (see for example [22]). On the other hand, for  $|\lambda| = 1$  the excluded unit roots are dense, and, moreover, one can find a dense set of  $\lambda$  on the unit circle which are not unit roots and for which the above series diverges [23]. But all the exceptional values can be very well approximated by roots of unity. If we require, however, that  $\lambda$  satisfies the infinitely many inequalities

$$(1.5) \quad |\lambda^q - 1|^{-1} \leq c_0 q^2 \quad \text{for} \quad q = 1, 2, \dots$$

then the series  $u$  converges in a neighborhood of  $\zeta = 0$ . This is the content of Siegel's theorem.

Siegel's original proof depends on delicate estimates which take into account that the number  $|\lambda^q - 1|^{-1}$  is usually much smaller than  $c_0 q^2$ . The proof which will be presented here is much cruder and therefore is applicable to the more difficult problems of celestial mechanics. We shall explain the method in a more general setting to bring out the important features. The detailed estimates for the proof of this theorem will be given in Section 3.

Here we merely consider the linearized equation for (1.3). Let  $u = w + \varepsilon v$  and differentiate (1.3) at  $\varepsilon = 0$  to get

$$(1.6) \quad v(\lambda\zeta) - f'(w(\zeta))v(\zeta).$$

If we could replace  $f(z)$  by its linear part  $\lambda z$  and  $w(\zeta)$  by  $\zeta$  then we are led to the equation

$$(1.7) \quad v(\lambda\zeta) - \lambda v(\zeta) = g(\zeta)$$

for a given power series  $g(\zeta)$  without linear and constant terms. This equation can be solved by power series expansion: If

$$g(\zeta) = \sum_{k>1} g_k \zeta^k$$

we find

$$v(\zeta) = \sum_{k>1} \frac{g_k}{\lambda^k - \lambda} \zeta^k.$$

If  $g$  converges in  $|\zeta| < \varrho$  then

$$|g_k| \leq c\varrho^{-k}$$

and by (1.5)

$$\left| \frac{g_k}{\lambda^k - \lambda} \right| \leq c_0 c\varrho^{-k} k^2.$$

Hence  $v(\zeta)$  converges in  $|\zeta| < \varrho$  also. The equation (1.7) corresponds to the linearized equation for  $w = \zeta$  and it seems hopeless to solve the corresponding equation in which the left-hand side of (1.7) is replaced by the expression (1.6). Therefore, in this case we have a situation in which the solution of the linearized equation (1) (of the introduction to Chapter III) is available only for  $w = \text{identity}$ .

## § 2. A construction for conjugacy problems.

The problem discussed in Section 1 can formally be written as

$$f \circ u = u \circ \Phi$$

or

$$u^{-1} \circ f \circ u = \Phi$$

where  $\Phi(\zeta) = \lambda\zeta$ . The circle  $\circ$  indicates composition of functions.

We introduce the functional

$$\mathcal{F}(f, u) = u^{-1} \circ f \circ u$$

and observe that it satisfies the important relations

$$(2.1) \quad \begin{cases} \mathcal{F}(f, u \circ v) = \mathcal{F}(\mathcal{F}(f, u), v) \\ \mathcal{F}(f, I) = f \text{ where } I \text{ denotes the identity map.} \end{cases}$$

The second argument  $u$  in  $\mathcal{F}(f, u)$  is to be considered as an element of a transformation group in the neighborhood of the identity. In our case we are dealing with the group of conformal mappings  $\zeta \rightarrow u(\zeta)$  with  $u(0) = 0$ ;  $u'(0) = 1$ .

Our problem is to solve the equation

$$(2.2) \quad \mathcal{F}(f, u) = \Phi$$

for  $u$  where  $f = \lambda z + \dots$  and  $\Phi = \lambda \zeta$  are given. We shall describe an iteration process which converges rapidly. The detailed estimates will be discussed later. We set  $u_0 = I$  and assume  $u_1, \dots, u_n$  have already been constructed. Then we set

$$(2.3) \quad u_{n+1} = u_n \circ v$$

where  $v = I + \hat{v}$  is to be found in such a manner that the functional equation

$$(2.4) \quad \mathcal{F}(f, u_n \circ v) = \Phi$$

holds — at least up to terms linear in  $\hat{v}$  and the error  $\mathcal{F}(f, u_n) - \Phi$ . We set

$$f_n = \mathcal{F}(f, u_n)$$

so that  $f_n - \Phi$  is small already. By (2.1) the equation takes the form

$$(2.5) \quad \mathcal{F}(f_n, v) = \Phi.$$

Expanding the left-hand side formally at the pair  $(\Phi, I)$  we obtain

$$\mathcal{F}(\Phi, I) + \mathcal{F}_f(\Phi, I)(f_n - \Phi) + \mathcal{F}_u(\Phi, I)\hat{v} = \Phi.$$

Here

$$\mathcal{F}_f(\Phi, I)g = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} (\mathcal{F}(\Phi + \varepsilon g, I) - \mathcal{F}(\Phi, I)) = g$$

by (2.1) and  $\mathcal{F}(\Phi, I) = \Phi$ . Therefore the above relation reduces to

$$(f_n - \Phi) + \mathcal{F}_u(\Phi, I)\hat{v} = 0$$

or

$$(2.6) \quad \mathcal{F}'(\Phi, I)\hat{v} = \Phi - f_n.$$

Here we replaced the notation  $\mathcal{F}_u$  by  $\mathcal{F}'$ :

$$\mathcal{F}'(Q, I) \hat{v} = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} (\mathcal{F}(\Phi, I + \varepsilon \hat{v}) - \mathcal{F}(\Phi, I)).$$

Therefore, if (2.6) can be solved for  $\hat{v}$  this equation together with (2.3) defines the next approximation. At least formally this process converges quadratically: If the error  $f_n - \Phi$  is of order  $\varepsilon_n$  in some norm, then from (2.6) also  $\hat{v} = v - I$  is of order  $\varepsilon_n$ . But since we determined  $\hat{v}$  in such a way that equation (2.5) is satisfied up to terms linear in  $f_n - \Phi$  and  $\hat{v}$ , the error in that equation and hence  $\varepsilon_{n+1}$  will be of order  $\varepsilon_n^2$ .

We illustrate this process with a simple example: Let  $A$  be real  $n$  matrix. The problem is to find  $(I - A)^{-1}$  by an iteration which does not require the inversion of a matrix at any step.

Denote

$$f = I - A$$

and  $u$  an arbitrary  $n$  by  $n$  matrix. If

$$\mathcal{F}(f, u) = f \cdot u$$

is the matrix product, the equation to be solved is

$$\mathcal{F}(f, u) = I.$$

This functional  $\mathcal{F}(f, u) = f \cdot u$  clearly satisfies the relations (2.1) and our construction yields

$$u_{n+1} = u_n \circ v$$

with

$$\mathcal{F}(I, I) v = I - \mathcal{F}(f, u_n)$$

or

$$\hat{v} = I - f \cdot u_n.$$

Setting

$$f u_n = I - A_n$$

we have

$$v = I + \hat{v} = I + A_n$$

and

$$A_{n+1} = I - (I - A_n) v = A_n^2.$$

Thus we find the explicit formula

$$A_n = A^{2^n}$$

and for the solution

$$u = (I - A)^{-1} = \prod_{n=0}^{\infty} (I + A^{2^n}).$$

This is the well-known Euler product which obviously converges quadratically if  $|A| < 1$  in some norm.

For the problem of Section 1 :

$$\mathcal{F} = u^{-1} \circ f \circ u$$

we find

$$\begin{aligned} \mathcal{F}'(\Phi, I) \hat{v} &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \{ (I + \varepsilon \hat{v})^{-1} \circ \Phi \circ (I + \varepsilon \hat{v}) - \Phi \} \\ (2.7) \qquad \qquad &= \lambda \hat{v} - \hat{v}(\lambda \xi). \end{aligned}$$

This operator is invertible as we have shown at the end of the previous section and we shall show now that the above construction converges to a solution of the center problem.

We emphasize again that in the above construction just the operator  $\mathcal{F}'(\Phi, I)$  has to be inverted and not  $\mathcal{F}'(f, u)$ . This is crucial for the success in small divisor problems in which small changes of the linear operator  $\mathcal{F}'$  may destroy invertibility, since the spectrum of  $\mathcal{F}'(\Phi, I)$  comes arbitrarily close to 0.

### § 3. Proof of Siegel's theorem.

After these motivations we give the details of the proof for Siegel's theorem.

We assume that the given mapping

$$z_1 = f(z) = \lambda z + \hat{f}(z)$$

is defined in a circle  $|z| < r$  and that

$$(3.1) \qquad |\hat{f}'| < \varepsilon \text{ for } |z| < r.$$

Since  $f$  does not contain constant or linear terms we can make  $\varepsilon$  arbitrarily small by choosing  $r$  sufficiently small. Furthermore, let  $\lambda$  satisfy the inequalities (1.5) and  $0 < |\lambda| \leq 1$ . The first step is to estimate the solution of the equation (1.7).



LEMMA 1: If  $g$  is analytic in  $|\zeta| < r$ , and satisfies  $|g| < \varepsilon$  there and  $g(0) = g'(0) = 0$ , then the function

$$v = \sum_{k=2}^{\infty} (\lambda^k - \lambda)^{-1} g_k \zeta^k$$

is analytic in  $|\zeta| < r$  and satisfies

$$|v| < 2c_0 \frac{\varepsilon}{\theta^3} \text{ for } |\zeta| < r(1 - \theta)$$

for  $0 < \theta < 1$ .

PROOF: By Cauchy's estimate we have  $|g_k| < \varepsilon r^{-k}$ . Hence

$$\begin{aligned} |v| &< \varepsilon c_0 \sum_{k=2}^{\infty} k^2 \left(\frac{\zeta}{r}\right)^k \\ &\leq \varepsilon c_0 \sum_{k=2}^{\infty} k^2 (1 - \theta)^k \leq \frac{2\varepsilon c_0}{\theta^3} \quad (1). \end{aligned}$$

According to the construction described in Section 2 we choose the transformation

$$z = v(\zeta) = \zeta + \hat{v}(\zeta)$$

by solving<sup>(2)</sup>

$$(3.2) \quad \hat{v}(\lambda\zeta) - \lambda\hat{v}(\zeta) = \hat{f}(\zeta).$$

Therefore applying Lemma 1 to  $g = \zeta\hat{f}'(\zeta)$  we find that

$$|\hat{v}'| < 2c_0 \frac{\varepsilon}{\theta^3} \text{ in } |\zeta| < r(1 - \theta).$$

This implies, with  $\hat{v}(0) = 0$  that

$$(3.3) \quad |\hat{v}| < \frac{2c_0 \varepsilon}{\theta^3} r.$$

---

<sup>(1)</sup> We use that for  $0 < x < 1$

$$\sum_{k=2}^{\infty} k^2 x^k \leq \sum_{k=2}^{\infty} k(k-1)x^k + \sum_{k=1}^{\infty} kx^k = \frac{x(1+x)}{(1-x)^3} \leq \frac{2}{(1-x)^3}.$$

<sup>(2)</sup> This equation corresponds to (2.6) as one sees from (2.7).

Secondly we investigate where  $v^{-1} \circ f \circ v$  is defined. For this purpose we prove

LEMMA 2: If

$$2c_0 \varepsilon < \theta^4 \text{ and } 0 < \theta < \frac{1}{4}.$$

Then the mapping  $z = v(\zeta) = \zeta + \hat{v}(\zeta)$  maps

$$|\zeta| < r(1 - 4\theta) \text{ into } |z| < r(1 - 3\theta).$$

Secondly, the image of the disc

$$|\zeta| < r(1 - \theta) \text{ covers the disc } |z| < r(1 - 2\theta).$$

PROOF: The first part follows immediately from (3.3): If  $|\zeta| < r(1 - 4\theta)$  then

$$\begin{aligned} |z| &\leq |\zeta| + |\hat{v}| < r\left(1 - 4\theta + \frac{2c_0 \varepsilon}{\theta^3}\right) \\ &< r(1 - 4\theta + \theta) = r(1 - 3\theta) \end{aligned}$$

by assumption of the lemma.

To prove the second part we have show that for  $|z| < r(1 - 2\theta)$  the equation

$$\zeta + \hat{v} = z$$

has a solution in  $|\zeta| < r(1 - \theta)$ . By Rouché's theorem it suffices to verify the inequality

$$|\hat{v}| \leq r\theta < |\zeta| - |z| \text{ for } |\zeta| = r(1 - \theta).$$

This is again a consequence of (3.3) and the assumption of the lemma.

LEMMA 3: If

$$(3.4) \quad 2c_0 \varepsilon < \theta^4 \text{ and } 0 < \varepsilon < \theta < \frac{1}{5}$$

then the mapping

$$\Phi = v^{-1} \circ f \circ v \text{ or } \zeta_1 = \Phi(\zeta)$$

is defined for

$$|\zeta| < r(1 - 5\theta) = \varrho.$$

Moreover, if we write this mapping in the form

$$\zeta_1 = \lambda\zeta + \widehat{\Phi}$$

we have

$$|\widehat{\Phi}'| < c_1 \frac{\varepsilon^2}{\theta^4} \text{ for } |\zeta| < \varrho,$$

where  $c_1 < 3c_0$ .

PROOF: By the Lemma 2 the disc  $|\zeta| < r(1 - 4\theta)$  is mapped by  $v$  into  $|z| < r(1 - 3\theta)$ . The function  $f = \lambda z + \widehat{f}$  is defined there and maps this disc — by (3.1) — into

$$|z| \leq r(1 - 3\theta) + r \cdot \varepsilon < r(1 - 2\theta).$$

Finally, by Lemma 2,  $v^{-1}$  is defined there and hence also  $\Phi$ .

To estimate  $\Phi$  we write the relation

$$v \circ \Phi = f \circ v$$

in terms of  $\widehat{v} = v - \zeta$ ,  $\widehat{f} = f - \lambda z$ ,  $\widehat{\Phi} = \Phi - \lambda\zeta$ . We find

$$\widehat{\Phi} + \widehat{v}(\lambda\zeta + \widehat{\Phi}) = \widehat{v}(\zeta) + \widehat{f}(v).$$

Since we chose  $\widehat{v}$  as a solution of (3.2) we get

$$\widehat{\Phi} = \widehat{v}(\lambda\zeta) - \widehat{v}(\lambda\zeta + \widehat{\Phi}) + \widehat{f}(v) - \widehat{f}(\zeta).$$

Estimating the right-hand side by the mean value theorem we find

$$|\widehat{v}(\lambda\zeta) - \widehat{v}(\lambda\zeta + \widehat{\Phi})| \leq \sup |v'| \sup |\widehat{\Phi}| \leq \frac{1}{5} \sup |\widehat{\Phi}|$$

where we used that by (3.4)

$$|v'| < \frac{2\varepsilon c_0}{\theta^3} < \theta < \frac{1}{5}.$$

Hence — by (3.3) —

$$\begin{aligned} \frac{4}{5} \sup |\widehat{\Phi}| &\leq |\widehat{f}(v) - \widehat{f}(\zeta)| \leq \sup |\widehat{f}'| |v - \zeta| \\ &< \varepsilon \frac{2c_0 \varepsilon}{\theta^3} r \end{aligned}$$

in  $|\zeta| < (1 - 4\theta)r$ . Shrinking the domain to  $|\zeta| < (1 - 5\theta)r$  Cauchy's estimate yields

$$|\widehat{\Phi}'| \leq \varepsilon \frac{5}{2} c_0 \frac{\varepsilon}{\theta^4}$$

which proves the lemma.

We have succeeded to transform the original mapping  $f(z)$  which satisfied (3.1) into a new one  $\Phi$  which is much closer to the linear mapping. We shall repeat this construction and show that the obtained sequence of mappings  $f_n$  converges. In fact, if  $f_n = \lambda z + \widehat{f}_n$  and

$$|\widehat{f}_n'(z)| \leq \varepsilon_n \text{ in } |z| < r_n$$

then Lemma 3 ensures that

$$f_{n+1} = v_n^{-1} \circ f_n \circ v_n$$

satisfies

$$(3.5) \quad |\widehat{f}_{n+1}'| \leq c_1 \frac{\varepsilon_n^2}{\theta_n^4} = \varepsilon_{n+1} \text{ in } |z| < r_{n+1}.$$

The radii  $r_n$  of the discs have to decrease with  $n$  and we choose

$$(3.6) \quad r_n = r(1 + 2^{-n})/2.$$

Then we define  $\theta = \theta_n$  by

$$\frac{r_{n+1}}{r_n} = 1 - 5\theta_n$$

so that

$$(3.7) \quad 5\theta_n = \frac{1}{(2^n + 1)2}.$$

The convergence will be ensured if we can show that the sequence  $\varepsilon_n$  defined by (3.5) tends to zero. From (3.5), (3.7) we find

$$\varepsilon_{n+1} \leq c_2^{n+1} \varepsilon_n^2, \quad n = 0, 1, \dots$$

Clearly, if  $\varepsilon_0$  is chosen small enough  $\varepsilon_n$  tends to zero. Indeed,

$$\varepsilon'_n = c_2^{n+2} \varepsilon_n$$

satisfies

$$\varepsilon'_{n+1} \leq \varepsilon_n'^2$$

i. e. for

$$\varepsilon'_0 = c_2^2 \varepsilon_0 < 1$$

we have convergence. We have to verify the validity of (3.4) for  $\theta = \theta_n$ ,  $\varepsilon = \varepsilon_n$  which is straightforward for sufficiently small  $\varepsilon_0$ .

Thus we have found a sequence of transformations  $v_0, v_1, \dots, v_n, \dots$  where  $v_n$  transforms the mapping  $f_n$  into  $f_{n+1}$ . Here  $f_0 = f$  is our given mapping. Hence

$$u_n = v_0 \circ v_1 \circ \dots \circ v_{n-1}$$

transforms  $f$  into

$$(3.8) \quad f_n = u_n^{-1} \circ f \circ u_n.$$

This mapping  $u_n$  is defined for

$$|\zeta| < r_{n-1}$$

as one verifies from Lemma 2. Namely  $v_{n-1}$  maps  $|\zeta| < r_{n-1}$  into  $|\zeta| < r_{n-2}$  etc. Moreover, by construction  $f_n$  is defined in the same disc. The  $r_n$  were chosen  $\geq r/2$  (see (3.6)) and it is easy to show that  $u_n$  converges uniformly in  $|\zeta| < r/2$ . For this purpose we consider the product

$$u'_n = \prod_{\nu=0}^{n-1} v'_\nu = \prod_{\nu=0}^{n-1} (1 + \hat{v}'_\nu)$$

where the derivative  $v'_\nu$  has to be evaluated at the point

$$v_{\nu+1} \circ \dots \circ v_{n-1}(\zeta).$$

The estimate before (3.3) ensures that

$$|\hat{v}'_\nu| \leq c_3^\nu \varepsilon_\nu$$

and thus the infinite product

$$\prod_{\nu=0}^{\infty} (1 + |\hat{v}'_\nu|) \leq c_4$$

converges uniformly in  $|\zeta| < r/2$ . This implies that

$$|u_{n+1} - u_n| \leq c_4 \sup |v_n(\zeta) - \zeta| \leq c_4 |\hat{v}_n| \rightarrow 0.$$

Thus  $u_n(\zeta) \rightarrow u(\zeta)$  and  $f_n(\zeta) \rightarrow \lambda\zeta$  and from (3.8)

$$u^{-1} \circ f \circ u(\zeta) = \lambda\zeta$$

which was to be shown. Since  $v_n(0) = 0$ ,  $v_n'(0) = 1$  we also have  $u(0) = 0$ ,  $u'(0) = 1$ . Thus the above construction succeeds if  $\varepsilon_0 = \sup_{|z| < r} |\hat{f}'|$  is chosen sufficiently small. This can be achieved by choosing  $r$  sufficiently small since  $f$  vanishes quadratically at  $z = 0$ . This completes the proof of Siegel's theorem.

§ 4. A theorem by N. Levinson.

We mention some examples of functionals  $\mathcal{F}(f, u)$  which satisfy the relations (2.1) and therefore qualify for the above approach. For example, let  $u = u(x)$  be a differentiable transformation of  $x \in E_n$  into  $E_n$  and  $\dot{u}'(x)$  its Jacobian matrix. Then

$$(4.1) \quad \mathcal{F}(f, u) = u'^{-1}(f \circ u)$$

satisfies (2.1). Indeed the above expresses the transformation law for a differential equation

$$\dot{x} = f(x)$$

under a transformation  $\dot{y} = u(x)$ .

If  $f$  denotes a mapping from one space  $X = E_n$  into another  $Y = E_m$  then

$$(4.2) \quad \mathcal{F}(f, u) = u_1^{-1} \circ f \circ u_2$$

expresses the transformation law under two automorphisms  $u_1$  of  $X$  and  $u_2$  of  $Y$ . The equation  $\mathcal{F}(f, u) = g$  expresses the equivalence of a mapping  $f$  and  $g$  from  $X$  to  $Y$  under appropriate coordinate changes.

In case  $X = Y$  the functional

$$(4.3) \quad \mathcal{F}(f, u) = u^{-1} \circ f \circ u$$

expresses the transformation law of a mapping of  $X$  into  $X$ . The functional discussed in Section 2 was of this type. Finally,

$$\mathcal{F}(f, u) = f \circ u$$

also satisfies our relation (2.1).

As an example for the latter functional we mention a theorem by N. Levinson [25] which belongs to the theory of several complex variables:

**THEOREM:** Let

$$f(z, w) = p_0(z, w) + w^{n+1} \hat{f}(z, w)$$

be a power series, convergent in  $|z| < \rho$ ,  $|w| < \sigma$  where  $p_0$  is a polynomial in  $w$  of degree  $\leq n$  with  $p_0(0, w) \equiv w^n$  and  $\hat{f}$  any power series. Then there exists a coordinate change

$$w = u(z, \omega) = \omega + \omega^2 \hat{u}(z, \omega)$$

such that

$$\Phi(z, \omega) = f(z, u(z, \omega))$$

is a polynomial of degree  $\leq n$  in  $\omega$ .

Introducing the functional

$$\mathcal{F}(f, u) = f(z, u(z, \omega))$$

we try to solve the equation

$$\mathcal{F}(f, u) = \omega^n \pmod{P_n}$$

where  $P_n$  represents the space of polynomials  $p$  of degree  $\leq n$  in  $\omega$  which vanish for  $z = 0$ .

Using the method described in Section 2 we reduce the problem to solving the linearized equation

$$(4.4) \quad \mathcal{F}'(\Phi, I)v = g \pmod{P_n}.$$

One finds easily that

$$\mathcal{F}'(\Phi, I) = \Phi_\omega(z, \omega)v$$

where  $\Phi_\omega$  is a polynomial of degree  $\leq n-1$  and  $\Phi_\omega(0, \omega) = n\omega^{n-1}$ . With  $v = \omega^2 \hat{u}$  the equation (4.4) reduces to

$$(\omega^2 \Phi_\omega) \cdot \hat{u} = g \pmod{P_n}.$$

This is a standard division problem and we choose a polynomial  $p(z, \omega) \in P_n$  in such a way that

$$p(z, \omega_\nu) = g(z, \omega_\nu), \quad \nu = 1, \dots, n + 1$$

at the  $n + 1$  roots of  $\omega^2 \Phi_\omega$ , which can be found uniquely by Lagrange's interpolation method. Then

$$\widehat{u} = \frac{g - p}{\omega^2 \Phi_\omega} = \frac{1}{2\pi i} \oint \frac{g(z, \lambda) d\lambda}{\lambda^2 \Phi_\omega(z, \lambda) (\lambda - \omega)}$$

gives the solution to (4.4), if the integration is taken over a circle containing all roots of  $\omega^2 \Phi_\omega$ .

The convergence proof can now be established with standard estimates. We refer to Levinson's paper for the details. It is noteworthy that the solution  $u(z, \omega)$  can also be found by comparison of coefficients, however, the majorant method by Cauchy does not seem to yield the convergence proof, while the iteration method described here succeeds.

### 5. Vector field on a torus and Kolmogorov's theorem.

As another application we discuss the following problem concerning vector fields on a torus: Let  $x = (x_1, \dots, x_n)$  and consider the vector field

$$(5.1) \quad \dot{x} = f(x)$$

where  $f(x)$  is an  $n$ -vector whose components have the period  $2\pi$  with respect to  $x_1, \dots, x_n$ . Therefore we can write (5.1) as a vector field on the torus which is obtained by identifying all points  $x$  whose coordinates differ by an integer multiple of  $2\pi$ .

The simplest model of such a differential equation is obtained if  $f(x)$  is independent of  $x$ :

$$(5.2) \quad \dot{x} = \omega.$$

It is well known that this flow is ergodic (with respect to the measure  $dx = dx_1, \dots, dx_n$ ) if and only if the components  $\omega_1, \dots, \omega_n$  of  $\omega$  are rationally independent.

We pose the question whether the flow (5.2) is structurally stable, i. e. whether small perturbations of (5.2) lead to a differential equation which



can be transformed into a (5.2) by a coordinate transformation

$$x = u(\xi) = \xi + \widehat{u}(\xi)$$

where  $u(\xi) - \xi$  have period  $2\pi$  in  $\xi_1, \dots, \xi_n$ , the components of  $\xi$ .

This is clearly not possible, since even a small change of the constants  $\omega$  leads to a flow

$$\dot{x} = \beta$$

which cannot be transformed into (5.2) unless  $\omega = \beta$ . Namely otherwise the transformation would transform the solution  $\xi = \beta t$  into a solution of (5.2), i. e.

$$\omega t + c = \beta t + \widehat{u}(\beta t).$$

Since  $\widehat{u}$  is bounded, being periodic in  $t$  it follows that  $\omega = \beta$ .

Therefore we shall admit changes of the vector  $f$  by a constant vector  $\lambda$  and shall try to determine it in such a manner that

$$\dot{x} = f(x) + \lambda$$

can be transformed into a system of the form (5.2). More precisely, let

$$(5.3) \quad \dot{x} = f(x, \varepsilon) = \omega + \varepsilon \widehat{f}(x, \varepsilon)$$

be a system where  $f(x, \varepsilon)$  is real analytic in  $x_1, \dots, x_n, \varepsilon$  and of period  $2\pi$  in  $x_\nu$ . Moreover, we shall require that the  $\omega_1, \dots, \omega_n$  are not only rationally independent but even satisfy the infinitely many inequalities

$$(5.4) \quad \left| \sum_{\nu=1}^n j_\nu \omega_\nu \right|^{-1} \leq c_0 \left( \sum_{\nu=1}^n |j_\nu| \right)^\tau$$

for all integers  $j_1, \dots, j_n$ , not all zero. If one chooses  $c_0$  large enough and  $\tau > n - 1$  then this condition is fulfilled for the majority of  $\omega$ , i. e. for almost all  $\omega$  such  $c_0$  exists.

**THEOREM 1:** Under the above conditions there exists a real analytic transformation

$$(5.5) \quad x = u(\xi, \varepsilon) = \xi + \widehat{u}(\xi, \varepsilon)$$

and a constant vector  $\lambda = \lambda(\varepsilon)$  with  $\lambda(0) = 0$  such that (5.5) transforms the

system

$$\dot{x} = \omega + \varepsilon \widehat{f}(x, \varepsilon) + \lambda(\varepsilon)$$

into

$$\dot{\xi} = \omega.$$

This theorem is due to V. I. Arnold (see [8]) who proved this result by the same method as described in Section 2. In the above form the statement does not appear to be useful as it refers to a modified system and not the given differential equation. However, this is to be considered as a partial step towards a more general theorem on differential equations in which sufficiently many parameters are available to achieve  $\lambda = 0$ . We now formulate such a theorem — which is due to Kolmogorov [6] and Arnold [10].

Consider a Hamiltonian system

$$(5.6) \quad \begin{cases} \dot{x} = H_y(x, y, \varepsilon) \\ \dot{y} = -H_x(x, y, \varepsilon). \end{cases}$$

with a real analytic Hamiltonian  $H(x, y, \varepsilon)$  of period  $2\pi$  in  $x_1, \dots, x_n$ . Moreover, assume that

$$H(x, y, 0) = \overset{\circ}{H}(y)$$

is independent of  $x$ . Hence, for  $\varepsilon = 0$  the system (5.6) takes the simple form

$$\dot{x} = \overset{\circ}{H}_y(y)$$

$$\dot{y} = 0$$

and the solutions are simply

$$x = \overset{\circ}{H}_y(\overset{\circ}{y})t + \overset{\circ}{x}; \quad y = \overset{\circ}{y}$$

where  $\overset{\circ}{x}, \overset{\circ}{y}$  are the initial values. The  $x$  components of the equations correspond to the system (5.3) discussed before. In order to have the parameters  $\overset{\circ}{y}$  available to adjust the vectors  $\overset{\circ}{H}_y(\overset{\circ}{y})$  (so as to achieve  $\lambda = 0$  in Theorem 1) we require that the Hessian is not zero:

$$\det \frac{\partial^2 H(y)}{\partial y \partial y} \neq 0.$$

**THEOREM 2:** Under the above hypothesis and for sufficiently small  $|\varepsilon|$  there exists a real analytic canonical transformation

$$x = u(\xi, \eta, \varepsilon)$$

$$y = v(\xi, \eta, \varepsilon)$$

such that  $u - \xi, v$  have period  $2\pi$  in  $\xi_1, \dots, \xi_n$ , that  $u = \xi, v = \eta$  for  $\varepsilon = 0$  and such that (5.6) is transformed into a system which reduces to

$$\dot{\xi} = \omega$$

$$\dot{\eta} = 0$$

for  $\eta = 0$ . Hence

$$x = u(\omega t + \overset{\circ}{\xi}, 0, \varepsilon); \quad y = v(\omega t + \overset{\circ}{\xi}, 0, \varepsilon)$$

represent a family of quasi-periodic solutions of the given unmodified system.

This result is of fundamental importance in the study of Hamiltonian systems and its applications to celestial mechanics (see Arnold [9]). We are not going into the proof of this result which can be found in [10] but rather discuss further the proof of Theorem 1 and its extension to differentiable vector fields.

### § 6. Proof of Theorem 1 (Analytic Case).

a) We turn to the proof of Theorem 1 as stated in the previous section. As mentioned before this result is not new but contained in Arnold's paper [8]: We discuss the proof in such a form that it also can be used for differential equations for which  $f(x)$  is merely differentiable and not analytic. However, we defer that case to the next section and assume now that in the given differential equation

$$\dot{x} = \omega + f(x)$$

$f(x)$  is a real analytic vector function which has period  $2\pi$  in  $x_1, \dots, x_n$ . We did not indicate the parameter dependence on  $\varepsilon$ , and replace it by a smallness condition.

The theorem in question asserts the existence of a real analytic function  $u(\xi)$  and a constant  $\lambda$  such that the differential equation

$$(6.1) \quad \dot{x} = \omega + f(x) + \lambda$$

is transformed by

$$(6.2) \quad x = u(\xi)$$

into the equation

$$(6.3) \quad \dot{\xi} = \omega.$$

Moreover,  $u(\xi) - \xi = \widehat{u}$  has period  $2\pi$  in  $\xi_1, \dots, \xi_n$ .

More precisely we shall show :

ADDITION TO THEOREM 1: There exists a positive constant  $c^*$  dependent on  $n, \tau, c_0$ , such that for

$$\varepsilon < h^{\sigma+1}/c^*, \quad \sigma = \tau + 1, \quad h < 1,$$

and for

$$|f(x)| < \varepsilon \quad \text{in} \quad |\text{Im } x| < h$$

one can find a constant vector  $\lambda$  in

$$|\lambda| < 2\varepsilon$$

and the desired transformation  $u(\xi)$  satisfying

$$h^{-1} |\widehat{u}| + |\widehat{u}'(\xi)| \leq c \frac{\varepsilon}{h^{\sigma+1}} < \frac{1}{2} \quad \text{in} \quad |\text{Im } \xi| < \frac{h}{2}$$

where  $c$  is independent of  $\varepsilon, h$ .

This statement assures then the existence of a solution of the partial differential equation

$$(6.4) \quad \widehat{u}_\xi \omega = f(\xi + \widehat{u}) + \lambda$$

on the torus. Here  $\widehat{u}_\xi$  denotes the Jacobian matrix  $\left(\frac{\partial u_\nu}{\partial \xi_\mu}\right)$  and  $\omega$  the vector with the components  $\omega_1, \dots, \omega_n$  on which we impose the irrationality (see (6.5) below). Clearly the constant  $\lambda$  has to be chosen in such a manner that the mean value of the right hand side of (6.4) vanishes.

To point out the subtlety of the problem we mention that one cannot expect a solution of an equation of the type (6.4) if  $f(\xi + \widehat{u})$  is replaced by a function  $f(\xi, \widehat{u})$  of period  $2\pi$  in  $\xi$ . In fact, even for a function linear in  $\widehat{u}$ , say  $f(\xi, \widehat{u}) = f_0(\xi) + c\widehat{u}$  one finds counter examples, no matter how small  $f_0, c$  are chosen. The reason for this phenomenon is the fact that

the spectrum of the operator

$$\sum_{\nu=1}^n \omega_{\nu} \frac{\partial}{\partial \xi_{\nu}}$$

acting on the space of functions on the torus contains infinitely many eigenvalues in any neighborhood of zero. Therefore (6.4) cannot be treated like any partial differential equation but the dependence of  $f$  on  $\xi + \widehat{u}$  alone must be taken into account. This, however, is equivalent to the fact that (6.4) represents a transformation law as discussed in Section 2.

We shall assume that  $\omega_1, \dots, \omega_n$  are numbers which are rationally independent and, moreover, satisfy the inequalities

$$(6.5) \quad |(k, \omega)|^{-1} \leq c_0 |k|^{\tau}$$

for all integers with  $|k| = \sum |k_{\nu}| > 0$ . Here  $\tau$  is some number  $> n - 1$  and  $c_0$  a positive constant. It is easily shown that every sphere of a radius  $r$  large compared to  $c_0^{-1}$  contains at least one such  $\omega$  provided  $\tau > n - 1$ . In fact the set of  $\omega$  which violates the above condition for any choice of  $c_0$  forms a set of measure zero if  $\tau > n - 1$ . This is easily verified by estimating the measure of this set.

b) We begin with two lemmata. The first one refers to the solvability of the linear partial differential equation

$$(6.6) \quad v_x \omega = g(x)$$

or in components

$$\sum_{\nu=1}^n \omega_{\nu} v_{\mu x_{\nu}} = g_{\mu}(x).$$

We require  $g, v$  to have period  $2\pi$  in  $x_1, \dots, x_n$ . Clearly, a necessary condition for the solvability of (6.6) is that the mean value  $[g]$  of  $g$  vanishes.

LEMMA 1: If  $g(x)$  is a real analytic vector function of mean value zero which is bounded in  $|\operatorname{Im} x_{\nu}| < h$  then there exists a real analytic solution  $v$  of period  $2\pi$ , provided (6.5) holds. Moreover,

$$\sup_{|\operatorname{Im} x| < h - \delta} |v| \leq c \delta^{-\sigma} \sup_{|\operatorname{Im} x| < h} |g|$$

for  $0 < \delta < h < 1$  and  $\sigma = \tau + 1 > n$ .

Here  $c$  denotes a positive constant dependent on  $\tau, n, c_0$  only<sup>(3)</sup>.

---

<sup>(3)</sup> In the following  $c$  will stand for different constants and will not be distinguished in each case.

PROOF: Using the Fourier expansion of  $g$  one finds the solution immediately as

$$v = \sum_{k \neq 0} \frac{\gamma_k}{(k, \omega)} e^{i(k, x)}$$

where  $\gamma_k$  are the Fourier coefficients of  $g$ . Since  $g$  is analytic the coefficients  $\gamma_k$  decay exponentially with  $|k|$ . Assuming  $\sup |g| \leq 1$  for  $|\operatorname{Im} x| \leq h$  we find

$$|\gamma_k| = \frac{1}{(2\pi)^n} \left| \int_T g(x) e^{-i(k, x)} dx \right| \leq e^{-|k|h}$$

by shifting the surface of integration to  $|\operatorname{Im} x_\nu| = \pm h$ . Therefore the above series converges and for  $|\operatorname{Im} x_\nu| \leq h - \delta$  can be estimated by

$$(6.7) \quad |v| \leq \sum_{k \neq 0} |(k, \omega)|^{-1} e^{-|k|\delta}.$$

Replacing the small divisors  $(k, \omega)$  according to the estimate (6.5) we find

$$|v| \leq c_0 \sum_{k \neq 0} |k|^\tau e^{-|k|\delta} \leq c_1 \delta^{-\tau-n}$$

which proves the lemma with  $\sigma = \tau + n$ .

The estimate stated in the lemma with  $\sigma = \tau + 1$  is more delicate and is based on the observation that only a few of the denominators  $(k, \omega)$  are small. This fact was used in Siegel's original proof (see [22], p. 168) of this theorem (Section 1) and also in Arnold's work ([9], Lemma 2, p. 30). We present the proof — for the sake of completeness — for our situation:

For the following we shall use the norm  $|k| = \max_\nu |k_\nu|$  which is equivalent to the previous norm, but is more appropriate for the following considerations. Let  $K(\nu, r)$  denote the set of vectors  $k \neq 0$  with integer coefficients satisfying

$$|k| = r \quad \text{and} \quad 2^\nu < |(k, \omega)|^{-1} \leq 2^{\nu+1}$$

and let  $N = N(\nu, r)$  denote the number of points in  $K(\nu, r)$ . We shall show that

$$N(\nu, r) \leq c_1 r^{n-1} 2^{-\frac{\nu}{\tau}(n-1)}$$

Again  $c_1$  is a positive constant depending on  $c_0, \tau, n$  only. Note that this estimate is particularly sharp for large  $\nu$ .

To prove this statement we note: If  $k, k'$  are different vectors of  $K(\nu, r)$  then

$$c_0^{-1} |k - k'|^{-\tau} \leq |(\omega, k - k')| \leq |(\omega, k)| + |(\omega, k')| \leq 2^{-\nu+1}.$$

Hence the distance

$$|k - k'| \geq (c_0^{-1} 2^{\nu-1})^{1/\tau} = 2\varrho_\nu,$$

is very large for large  $\nu$ . We define  $\varrho_\nu$  by the above relation and note that from  $|k - k'| \leq 2r$  we have

$$(6.8) \quad \varrho_\nu \leq r.$$

If we surround each  $k \in K(\nu, r)$  by a cube

$$\mathcal{C}_k: |x - k| < \varrho_\nu,$$

then these cubes are disjoint. Intersecting the cubes  $\mathcal{C}_k$  with  $|x| = r$  we get disjoint  $n - 1$  dimensional sets of  $(n - 1)$  dimensional volume  $\geq \varrho_\nu^{n-1}$ . Since the  $n - 1$  dimensional volume of  $|x| = r$  is  $2n(2r)^{n-1}$  we have

$$N(\nu, r) \leq \frac{2n(2r)^{n-1}}{\varrho_\nu^{n-1}} \leq c_1 r^{n-1} 2^{-\frac{\nu}{\tau}(n-1)}$$

which proves the stated inequality.

To finish the proof of the lemma we form

$$\sum_{K(\nu, r)} |(k, \omega)|^{-1} \leq 2^{\nu+1} N(\nu, r) \leq 2c_1 r^{n-1} 2^{\left(1 - \frac{n-1}{\tau}\right)\nu}.$$

Note that the last exponential is positive for  $\nu > 0$ . Therefore adding over all  $\nu$  for which  $K(\nu, r)$  is not empty we get

$$\sum_{|k|=r} |(k, \omega)|^{-1} \leq c_2 r^{n-1} 2^{\left(1 - \frac{n-1}{\tau}\right)\nu^*},$$

where  $\nu^*$  is the greatest occurring  $\nu$  for which  $K(\nu, r) \neq \emptyset$ . The relation (6.8) provides us with an estimate for  $\nu^*$  and yields

$$2^{\nu^*/\tau} \leq c_3 r.$$

Hence

$$\sum_{|k|=r} |(k, \omega)|^{-1} \leq c_4 r^{n-1} r^{\tau-(n-1)} = c_4 r^\tau.$$

Finally, (6.7) is estimated by

$$|v| \leq \sum_{r=1}^{\infty} \sum_{|k|=r} |(k, \omega)|^{-1} e^{-r\delta} \leq c_4 \sum_{r=1}^{\infty} r^r e^{-r\delta} \leq c_5 \delta^{-r-1},$$

which was to be proven.

c) To prove Theorem 1 and its addition we consider a family of differential equations

$$(6.9) \quad \dot{x} = a + f(x, a)$$

which depends on a parameter  $a$  (analytically) which varies in a complex neighborhood of  $\omega$ . We shall find not only a transformation of the variables  $x$

$$(6.10) \quad x = u(\xi, \alpha)$$

but also a transformation

$$(6.11) \quad a = w(\alpha)$$

of the parameters in such a manner that in an appropriate domain the transformed equation

$$(6.12) \quad \dot{\xi} = \alpha + \Phi(\xi, \alpha)$$

possesses a function  $\Phi$  which is much smaller than  $f$ . Repeating this process we shall construct the solution asserted in Theorem 1.

To make the estimates precise we require the following inequalities

$$(6.13) \quad |f(x, a)| < \varepsilon \quad \text{for} \quad |\operatorname{Im} x| < s; \quad |a - \omega| < 2\varepsilon,$$

with appropriate positive numbers  $\varepsilon, s < 1$ . All estimates are considered in the complex domain.

LEMMA 2. Assume that  $f$  is real analytic in

$$\mathcal{D}: |\operatorname{Im} x| < s, \quad |a - \omega| < 2\varepsilon,$$

and satisfies (6.13). Let  $s_+$  be a positive number  $s_+ < s$  and let

$$\frac{\varepsilon}{(s - s_+)^{\sigma+1}}$$



be sufficiently small and set

$$\varepsilon_+ = c \frac{\varepsilon^2}{(s - s_+)^{\sigma+1}}$$

with an appropriate positive constant  $c$ .

Then there exists a transformation

$$\mathcal{U}: \begin{aligned} x &= u(\xi, \alpha) \\ a &= w(\alpha) \end{aligned}$$

which maps

$$\mathcal{D}_+: |\operatorname{Im} \xi| < s_+, \quad |\alpha - \omega| < 2\varepsilon_+$$

into  $\mathcal{D}$  and is real analytic there. Moreover,

$$s^{-1} |u - \xi|, \quad |u_\xi - I| < \frac{c\varepsilon}{(s - s_+)^{\sigma+1}}, \quad |w - \alpha| < \varepsilon,$$

and the transformed equation (6.12) satisfies

$$|\Phi| < \varepsilon_+ \quad \text{in} \quad \mathcal{D}_+.$$

PROOF: We define  $u(\xi, \alpha) = \xi + \widehat{u}$  by the equation

$$(6.14) \quad \widehat{u}_\xi \omega = f(\xi, a) - [f(\xi, a)]; \quad [\widehat{u}] = 0.$$

By Lemma 1 this equation has a real analytic solution which can be estimated by

$$|\widehat{u}| \leq c \frac{\varepsilon}{(s - s_+)^{\sigma}} \quad \text{in} \quad |\operatorname{Im} \xi| < s - \frac{s - s_+}{2},$$

and by Cauchy's estimate

$$|\widehat{u}_\xi| \leq c \frac{\varepsilon}{(s - s_+)^{\sigma+1}} \quad \text{in} \quad |\operatorname{Im} \xi| < s_+.$$

Secondly we define  $a = w(\alpha)$  in (6.11) implicitly by

$$(6.15) \quad \alpha = a + [f(\xi, a)].$$

That this equation can be solved for  $a$  if  $|\alpha - \omega| < 2\varepsilon_+$  follows immedia-

tely from an index argument using

$$|\alpha - \omega - [f]| \leq 2\varepsilon_+ + \varepsilon < 2\varepsilon^{(4)}.$$

Therefore the degree of our mapping in  $|a - \omega| \leq 2\varepsilon$  with respect to  $\alpha$  is 1. Hence  $a = w(\alpha)$  exists. Moreover, one verifies that the Jacobian of the mapping does not vanish so that  $w(\alpha)$  is analytic in  $|\alpha - \omega| < 2\varepsilon_+$ . The stated inequality for  $|w - \alpha|$  follows from (6.15).

The above estimates ensure that the transformation  $\mathcal{U}$  defined by (6.14), (6.15) maps  $\mathcal{D}_+$  into  $\mathcal{D}$ , since

$$|\operatorname{Im} x| \leq |\operatorname{Im} \xi| + |u| \leq s_+ + c \frac{\varepsilon}{(s - s_+)^\sigma} \leq s_+ + (s - s_+) = s.$$

Here we used the smallness assumption of Lemma 2.

Having defined  $u(\xi, \alpha)$ ;  $w(\alpha)$  we estimate  $\Phi$ . The transformation formula gives

$$u_\xi(\alpha + \Phi) = a + f(\xi + \widehat{u}, a).$$

Subtracting (6.14) and (6.15) from this relation we find

$$\Phi + \widehat{u}_\xi \Phi = f(\xi + \widehat{u}, a) - f(\xi, a) - \widehat{u}_\xi(\alpha - \omega).$$

Since  $|\widehat{u}_\xi| < 1/2$  we can estimate  $\Phi$  in  $|\operatorname{Im} \xi| < s_+$ ,  $|\alpha - \omega| < \varepsilon_+$  by

$$\frac{1}{2} |\Phi| \leq \sup |f'| \cdot |\widehat{u}| + |\widehat{u}_\xi| |\alpha - \omega|.$$

Using Cauchy's estimate we find

$$\begin{aligned} |\Phi| &\leq c \left\{ \frac{\varepsilon}{s - s_+} \frac{\varepsilon}{(s - s_+)^\sigma} + \frac{\varepsilon}{(s - s_+)^{\sigma+1}} \varepsilon_+ \right\} \\ &\leq c' \frac{\varepsilon^2}{(s - s_+)^{\sigma+1}}, \end{aligned}$$

which finishes the proof.

---

(4) We used that for sufficiently small  $\frac{\varepsilon}{(s - s_+)^{\sigma+1}}$  we have  $2\varepsilon_+ < \varepsilon$ .

d) CONVERGENCE PROOF. To prove Theorem 1 (of Section 5) and its addition (see beginning of this section) we apply Lemma 2 repeatedly. We start with the prescribed family of differential equations given by

$$f_0 = f(x) + a \quad (\text{where } a = \omega + \lambda)$$

and transform it by

$$\mathcal{U}_0: \begin{cases} x = u(\xi, \alpha) \\ a = w(\xi, \alpha) \end{cases}$$

into the new family

$$f_1 = f_1(\xi, \alpha) = \mathcal{F}(f_0, \mathcal{U}_0).$$

Here  $\mathcal{F}(f, \mathcal{U})$  represents the transformation law corresponding to differential equations, i. e.

$$\mathcal{F}(f_0, \mathcal{U}_0) = (u')^{-1} f(u(\xi, \alpha), w(\alpha)).$$

$f_1$ , in turn, is transformed by a transformation  $\mathcal{U}_1$  into

$$f_2 = \mathcal{F}(f_1, \mathcal{U}_1) = \mathcal{F}(f_0, \mathcal{U}_0 \circ \mathcal{U}_1)$$

etc. We shall show that the composition of the transformations  $\mathcal{U}_0, \mathcal{U}_1, \dots$ , are defined in appropriate domains and converge. In particular we shall show that

$$\left. \begin{array}{l} \mathcal{U}_1 \circ \mathcal{U}_2 \circ \dots \circ \mathcal{U}_k \rightarrow \mathcal{U}^* \\ f_k \rightarrow 0 \end{array} \right\} \text{for } \alpha = 0, |\operatorname{Im} \xi| < h/2.$$

Writing the coordinate transformation  $\mathcal{U}^*$  in the form

$$x = u^*(\xi, 0)$$

$$a = w^*(0)$$

we see that the first line represents the desired coordinate transformation while the second specifies the value of  $a$ , i. e.  $\lambda = a - \omega = w^*(0) - \omega$ .

We proceed to define  $\mathcal{U}_0, \mathcal{U}_1, \dots$ , inductively. Assume that  $\mathcal{U}_0, \mathcal{U}_1, \dots, \mathcal{U}_{k-1}$  have been defined already and

$$f_k = \mathcal{F}(f_0, \mathcal{U}_0 \circ \mathcal{U}_1 \circ \dots \circ \mathcal{U}_{k-1});$$

then we use the construction of Lemma 2 to define  $\mathcal{U}_k$  and

$$f_{k+1} = \mathcal{F}(f_k, \mathcal{U}_k) = \mathcal{F}(f_0, \mathcal{U}_0 \circ \mathcal{U}_1 \circ \dots \circ \mathcal{U}_k).$$

Inductively we shall verify the following estimates.

We define the domains

$$\mathcal{D}_k: |\operatorname{Im} x| < s_k, \quad |a - \omega| < 2\varepsilon_k,$$

where we set

$$(6.16) \quad \begin{cases} s_k = \frac{h}{2}(1 + 2^{-k}), & k = 0, 1, 2, \dots, \\ \varepsilon_k = c_1^k h^{-\sigma-1} \varepsilon_{k-1}^2, & k = 1, 2, \dots, \\ \varepsilon_0 = \varepsilon, \end{cases}$$

with an appropriate large constant  $c_1$  depending on  $c_0, \tau, \sigma, n$  only.

Then

$\alpha)$   $f_k$  is defined and analytic in  $\mathcal{D}_k$  and satisfies

$$|f_k - a| < \varepsilon_k$$

$\beta)$   $\mathcal{U}_k$  is defined and analytic in  $\mathcal{D}_{k+1}$  mapping  $\mathcal{D}_{k+1}$  into  $\mathcal{D}_k$ .

Moreover,

$$|w_k - \alpha| < \varepsilon_k, \quad h^{-1}|u_k - \xi|, \quad |u'_k - I| < c^{k+1} h^{-\sigma-1} \varepsilon_k.$$

We verify the statement for  $k = 0$ : In this case  $\alpha)$  follows from the hypothesis of the addition to Theorem 1 since

$$|f_0 - a| = |f(x)| < \varepsilon = \varepsilon_0 \text{ in } \mathcal{D}_0.$$

Lemma 2 ensures the existence of  $\mathcal{U}_0$  defined in  $\mathcal{D}_1$ , if we choose  $s = s_0$ ,  $s_+ = s_1$ .

Assume that the above statement has been verified for  $k = 0, 1, \dots, l$ . Then we can apply Lemma 2 to  $f = f_l$  in  $\mathcal{D} = \mathcal{D}_l$ , and find  $\mathcal{U} = \mathcal{U}_l$  in  $\mathcal{D}_{l+1}$  by the same lemma. The statements  $\alpha), \beta)$  follow directly from those of Lemma 2.

Having established the statements  $\alpha), \beta)$  for

$$f_k = \mathcal{F}(f_0, \mathcal{U}_0 \circ \mathcal{U}_1 \circ \dots \circ \mathcal{U}_{k-1})$$

we discuss now the limit process  $k \rightarrow \infty$ . Note that the transformation

$$\mathcal{V}_k = \mathcal{U}_0 \circ \mathcal{U}_1 \circ \dots \circ \mathcal{U}_{k-1}$$

is defined for a shrinking domain  $\mathcal{D}_k$ . We restrict this transformation to  $\mathcal{D} = \bigcap_{k \geq 0} \mathcal{D}_k$  which is given by

$$\mathcal{D}: |\operatorname{Im} \xi| < \frac{h}{2}, \quad \alpha = 0.$$

$\mathcal{V}_k$  maps  $\mathcal{D}$  into  $\mathcal{D}_0$  by construction.

To show finally, that  $\mathcal{V}_k$  converges in  $\mathcal{D}$  we write  $\mathcal{V}_k$  in components

$$\mathcal{V}_k: \begin{cases} x = v_k(\xi, \alpha) \\ a = \tilde{w}_k(\alpha). \end{cases}$$

By construction we have

$$|v_{k+1} - v_k| < c^{k+1} h^{-\sigma} \varepsilon_k; \quad |\tilde{w}_{k+1} - \tilde{w}_k| < \varepsilon_k,$$

and since

$$\sum_k c^{k+1} \varepsilon_k$$

converges by (6.16) (for sufficiently small  $\varepsilon/h^{\sigma+1}$ ) it is clear that  $\lim_k v_k$  exists and is analytic in  $|\operatorname{Im} \xi| < h/2$ . We find for  $\alpha = 0$

$$|\lim_{k \rightarrow \infty} w_k| < \sum \varepsilon_k \leq 2 \varepsilon_0 = 2 \varepsilon$$

for sufficiently small  $\varepsilon h^{-\sigma-1}$  so that  $\lim_{k \rightarrow \infty} \mathcal{V}_k = \mathcal{V}^*$ , or in components

$$\mathcal{V}^*: \begin{cases} x = v^*(\xi) \\ a = w^* \end{cases}$$

satisfies

$$|w^*| < 2 \varepsilon$$

and

$$|v^*(\xi) - \xi| < c \frac{\varepsilon}{h^\sigma}.$$

Similarly one verifies

$$|v^{*'}(\xi) - I| < c \frac{\varepsilon}{h^{\sigma+1}}.$$

This completes the proof of Theorem 1 (and the addition to it) if one sets  $\lambda = v^* + \omega$  and  $u = v^*(\xi)$ , except for the analytic dependence on the parameter — which was also called  $\varepsilon$  in Theorem 1. This can easily be taken care of by allowing all functions above to be analytic functions of this

additional parameter  $\varepsilon$  in a fixed domain  $|\varepsilon| < \mu$ . All approximations  $u_k, w_k$  will depend analytically on this parameter in one and the same domain  $|\varepsilon| < \mu$  and since the convergence is uniform the limit function will also depend analytically on the parameter.

### § 7. Vector field on a torus (differentiable case).

a) We use this opportunity to formulate and prove Theorem 1 on torus flows also in the differentiable case. Originally, in the proof of a similar result on invariant curves of area preserving annulus mappings the author treated the differentiable case by use of a particular smoothing operator which approximates functions in  $\mathcal{C}^r$  by  $\mathcal{C}^\infty$  functions where the error becomes extremely small if  $r$  is large. In the theory of approximations such results are well known, such as the theorem of Jackson which ensures that a function in  $\mathcal{C}^r$  can be approximated by trigonometrical polynomials of degree  $< N$  with an error  $< cN^{-r}$  (see N. I. Achieser, Chapter V [26]). We shall show that such approximation techniques can be used successfully to reduce the differentiable case to the analytic one. This approach follows the ideas of Bernstein who characterized the differentiable functions by their approximation properties by analytic functions. In a similar manner we approximate here the functional equation in the differentiable case by an analytic one. Aside from the interest per se this approach yields a result which requires a reasonable number of derivatives while our previous method [4''] was extremely wasteful in this respect.

Let  $\mathcal{C}^l$  denote the class of vector functions  $f(x)$  which have period  $2\pi$  in  $x_1, \dots, x_n$ , and continuous derivatives up to order  $l$ . For noninteger  $l > 0$  we require that the derivatives of order  $[l]$ <sup>(5)</sup> be Hölder continuous with exponent  $l - [l]$ . We denote by  $|f|_l$  the maximum of all derivatives of order  $\leq l$  if  $l$  is an integer. For noninteger  $l$  we add the Hölder constant to this expression.

We are going to prove the following

**THEOREM 3:** <sup>(6)</sup> Let  $l > \sigma + 1 = \tau + 2 > n + 1$  and let the vector  $\omega = (\omega_1, \dots, \omega_n)$  satisfy (5.4). Then there exists a positive constant  $\delta_0$  depen-

---

<sup>(5)</sup>  $[l]$  denotes the largest integer  $\leq l$ .

<sup>(6)</sup> Added in proof: In the meantime A. M. Samolenko studied the same problem in Ukrainskii Math. Journal, vol. 16, N. 6, 1964, pp. 769-782. He makes use of a smoothing operator as described above while we approximate the problem by an analytic one. This leads also to considerably milder differentiability requirements: Instead of  $l = 1 + 32(n + 2)$  we need only  $l > n + 1$ .

ding on  $c_0, \tau, l, n$  only such that for

$$|f(x) - \omega|_l < \delta < \delta_0$$

there exists a constant  $\lambda$  and a coordinate transformation

$$x = \xi + \widehat{u}(\xi)$$

which maps  $\dot{x} = f(x) + \lambda$  into

$$\dot{\xi} = \omega.$$

Here  $\widehat{u}$  belongs to  $\mathcal{C}^l$  (more precisely to  $\mathcal{C}^{l-\sigma}$  if  $l - \sigma$  is not an integer and satisfies

$$|\widehat{u}|_1 \leq c \delta; \quad |\lambda| < c \delta$$

with some constant  $c$  depending on  $c_0, \tau, l, n$  only.

Notice that  $u$  possesses  $\sigma$  derivatives less than  $f$ ; this is precisely the same loss of derivatives as one has for the linear partial differential equation (6.6). Therefore the drop of differentiability for the linearized equations is the same as for the corresponding nonlinear ones!

b) Before we prove Theorem 3 we shall show how one can approximate a function  $f(x) \in \mathcal{C}^l$  by analytic functions. This is a standard result of approximation theory but we indicate the proof.

**LEMMA 1.** Any function  $f(x) \in \mathcal{C}^l$  can be represented as

$$f = \lim_{k \rightarrow \infty} f_k(x), \quad \text{for real } x,$$

where  $f_k(x)$  is real analytic satisfying

$$(7.1) \quad |f_k(x) - f_{k+1}(x)| \leq A h_k^l \quad \text{for} \quad |\operatorname{Im} x| < h_k = 4^{-k}, \\ k = 0, 1, \dots,$$

and  $f_0 = 0$  with  $A \leq c |f|_l$ . Here  $c$  depends on  $l, n$  only. Conversely, if a real function  $f$  admits such an approximation by analytic ones then  $f \in \mathcal{C}^l$  provided  $l$  is not an integer and  $|f|_l \leq c' A$  with another constant  $c'$ .

**PROOF:** By Jackson's Theorem there exists a trigonometrical polynomial  $p_N(x)$  of degree  $< N$  such that

$$|f(x) - p_N(x)| < \frac{c_1}{N^l} |f|_l.$$

Such a trigonometric polynomial can be provided by convolution operators (see Achieser) for arbitrary dimensions  $n$ . We set  $N = 4^{-k}$  and  $f_k = p_N(x)$  for  $k = 1, \dots$ , and  $f_0 = 0$ . Then  $f_k$  is an entire function and for real  $x$

$$|f_k - f_{k+1}| \leq |f_k - f| + |f_{k+1} - f| \leq 2 c_1 h_k^l |f|_l.$$

Using a well known theorem by S. Bernstein we find that all derivatives of the trigonometric polynomial  $f_k - f_{k+1}$  can be estimated by

$$\begin{aligned} |f_k - f_{k+1}|_s &\leq (4^{k+1})^s \cdot 2 c_1 h_k^l |f|_l \\ &\leq c_2 h_k^{l-s} |f|_l. \end{aligned}$$

To estimate  $f_k - f_{k+1}$  in the complex domain, say for  $|\operatorname{Im} x| < h_k$ , we use Taylor's expansion to get

$$|f_k - f_{k+1}| \leq c_2 h_k^l \left( \sum_{s=0}^{\infty} \frac{1}{s!} \right)^n = c_2 |f|_l e^n h_k^l.$$

This proves the first part.

The second part follows from Cauchy's estimate but will not be proven here.

c) We shall make a slight extension of Theorem 1, § 5 (or its addition, § 6): That statement referred to the differential equation (6.1) where  $\lambda$  was a constant. We shall now require that this modifying term be of the form  $p(x)\lambda$  where the matrix  $p(x)$  is real analytic in  $|\operatorname{Im} x| < h$  and satisfies

$$(7.2) \quad |\lambda| \leq |p(x)\lambda| \leq 3|\lambda| \quad \text{for all vectors } \lambda.$$

Then one can show with the smallness assumptions of p. 19 that the system

$$(7.3) \quad \dot{x} = \omega + \hat{f}(x) + p(x)\lambda$$

can be transformed into (6.3) for some appropriate constant  $\lambda$  in  $|\lambda| < 2\varepsilon$  (7).

(7) For the proof we remark that a transformation

$$x \rightarrow x + \lambda u(x)$$

$$\lambda \rightarrow \lambda [p]$$

where  $u$  solves

$$u_x \omega = p(x) - [p]$$

reduces (7.3) to



d) To prove Theorem 3 apply Lemma 1 to the given  $f(x) - \omega \varepsilon \mathcal{C}^l$  and construct  $f_k(x)$  which are analytic in  $|\operatorname{Im} x| < 4 h_k = 4^{1-k}$  and satisfy (7.1) and  $f_0 = \omega$ . The plan is to use Theorem 1 (in its extended form) to prove the existence of constants  $\mu_k$  and a coordinate transformation  $\mathcal{U}_k$  which transforms  $f_k + \mu_k$  into  $\xi = \omega$ . In fact we claim:

For  $k = 0, 1, \dots$ , there exists a real analytic transformation

$$\mathcal{U}_k : x = U_k(\xi)$$

and a constant  $\mu_k$  such that

$$(7.3) \quad \mathcal{F}(f_k + \mu_k, \mathcal{U}_k) = \omega.$$

Moreover,  $\mathcal{U}_k$  maps the strip  $\mathcal{D}_{k+1} : |\operatorname{Im} x| < 4^{-k-1}$  into  $\mathcal{D}_k$  and satisfies

$$(7.4) \quad |U_k - \xi|_1 \leq c \delta.$$

for real  $\xi$  and

$$(7.4') \quad |\mu_{k+1} - \mu_k| < 8 c \delta \cdot 4^{-kl}.$$

We prove this statement by induction. For  $k = 0$  we can clearly take  $\mu_0 = 0$ ,  $U_0(\xi) = \xi$  since  $f_0 = \omega$ . We assume that the statement has been proven in the form as it stands and shall establish it for  $k$  replaced by  $k + 1$ .

We consider  $f_{k+1} + \mu_k + 2\lambda$  for a variable  $\lambda$  and subject it to the transformation  $\mathcal{U}_k$  obtaining

$$(7.5) \quad \mathcal{F}(f_{k+1} + \mu_k + 2\lambda, \mathcal{U}_k) = \Phi + 2 U_k'^{-1} \lambda$$

by which  $\Phi$  is defined. Comparing this equation with (7.3) we find

$$\begin{aligned} |\Phi - \omega| &= |\mathcal{F}(f_{k+1} + \mu_k, \mathcal{U}_k) - \mathcal{F}(f_k + \mu_k, \mathcal{U}_k)| \\ &\leq |U_k'^{-1}| |f_{k+1} - f_k| \leq |U_k'^{-1}| c h_k^l \delta \end{aligned}$$

---


$$x = \omega + \hat{f}_1(x) + \mathcal{O}(\lambda^2) + \lambda.$$

Therefore, in  $|\lambda| < 2\varepsilon$  we can estimate  $\mathcal{O}(\lambda^2)$  by  $\varepsilon$  if  $\varepsilon$  is sufficiently small. In this manner the above statement is reduced to Theorem 1, § 5.

from Lemma 1. If  $\delta$  is chosen sufficiently small we infer from (7.4)

$$(7.6) \quad |\Phi - \omega| < 2c h_k^l \delta \quad \text{in} \quad |\text{Im } x| < h_k.$$

Now we apply Theorem 1 to

$$\dot{x} = \overline{\Phi}(x) + 2 U_k'^{-1} \cdot \lambda$$

and use that

$$\mathbb{F} \leq |2 U_k'^{-1}| \leq 3$$

for sufficiently small. The estimate (7.6) assures that with

$$\varepsilon = 2c h_k^l \delta; \quad h = h_{k+1}$$

the quantity

$$\frac{\varepsilon}{h_{k+1}^{\sigma+1}} \leq 4^{\sigma+2} c \delta$$

is sufficiently small if  $\delta$  is taken small enough. Hence there exists a  $\lambda$  in

$$(7.7) \quad |\lambda| < 2\varepsilon = 4c h_k^l \delta$$

and a coordinate transformation  $x = u_{k+1}(\xi)$  taking  $|\text{Im } x| < \frac{1}{2} h_{k+1}$  into  $\mathcal{D}_{k+1}$  such that

$$\mathcal{F}(\Phi + 2 U_k'^{-1} \lambda, u_{k+1}) = \omega$$

or equivalently with (7.5) (using (2.1))

$$\mathcal{F}(f_{k+1} + \mu_k + 2\lambda, \mathcal{U}_k \circ u_{k+1}) = \omega.$$

Setting

$$\mathcal{U}_{k+1} = \mathcal{U}_k \circ u_{k+1},$$

$$\mu_{k+1} = \mu_k + 2\lambda,$$

we have

$$\mathcal{F}(f_{k+1} + \mu_{k+1}, \mathcal{U}_{k+1}) = \omega,$$

as was to be shown.

It remains to verify the various estimates. Since  $u_{k+1}$  takes

$$(7.8) \quad |\text{Im } \xi| < \frac{1}{2} h_{k+1}$$

into  $\mathcal{D}_{k+1}$  and  $\mathcal{U}_k$  takes  $\mathcal{D}_{k+1}$  into  $\mathcal{D}_k$  it is clear that

$$\mathcal{U}_{k+1} = \mathcal{U}_k \circ u_{k+1}$$

maps (7.8) into  $\mathcal{D}_k$ . This is not sufficient for the induction and will be improved upon below.

From the addition to Theorem 1 (§ 6) p. 19 we have the estimate

$$(7.9) \quad h^{-1} |u_{k+1} - \xi|_0 + |u_{k+1} - \xi|_1 \leq c \frac{\varepsilon_k}{h_{k+1}^{\sigma+1}} \leq c \delta h_k^\alpha$$

with  $\alpha = l - \sigma - 1 > 0$ . Therefore, representing  $\mathcal{U}_{k+1}$  in the form

$$\mathcal{U}_{k+1} = u_1 \circ u_2 \circ \dots \circ u_{k+1}$$

we find from the chain rule

$$(7.10) \quad |U_{k+1} - \xi|_1 \leq c' \delta \sum h_k^\alpha \leq c'' \delta$$

if  $\delta$  is chosen small enough. This ensures (7.4) in the domain (7.8).

We shall show now that  $\mathcal{U}_{k+1}$  maps  $\mathcal{D}_{k+2}$  into  $\mathcal{D}_{k+1}$ . For this purpose we observe that  $U(\xi) = U_{k+1}(\xi)$  is real analytic, hence with  $\operatorname{Re} \xi = \varrho$  we have

$$|\operatorname{Im} U(\xi)| = |\operatorname{Im}(U(\xi) - U(\varrho))| \leq |U'| \cdot |\operatorname{Im} \xi|.$$

By (7.10) we can assume that  $|U'| \leq 2$  and therefore

$$|\operatorname{Im} U(\xi)| \leq 2 |\operatorname{Im} \xi| < h_{k+1} \quad \text{for} \quad |\operatorname{Im} \xi| < h_{k+2}.$$

Hence  $\mathcal{U}$  maps  $\mathcal{D}_{k+2}$  into  $\mathcal{D}_{k+1}$  as we wanted to show.

Finally the relation (7.7) implies the estimate (7.4') for  $|\mu_{k+1} - \mu_k|$  which completes the induction.

From (7.9) one also deduces the convergence of the  $U_k$  for  $k \rightarrow \infty$  if  $\xi$  is real. Namely, for  $\xi \in \mathcal{D}_{k+2}$  we have from (7.9)

$$|U_{k+1} - U_k| \leq |U_k \circ u_{k+1} - U_k| \leq |U_k| |u_{k+1} - \xi| \leq 2 c \delta h_k^{\alpha+1}.$$

Hence  $\mathcal{U}_k$  certainly converges for real  $\xi$ . Moreover, by the second part of Lemma 1 it follows that the limit function

$$\lim_{k \rightarrow \infty} U_k = U^*$$

belongs to  $\mathcal{C}^{1+\alpha} = \mathcal{C}^{l-\sigma}$  if  $\alpha$  is not an integer and

$$|U^* - \xi|_{1+\alpha} \leq c_1 \delta.$$

Setting

$$\lim_{k \rightarrow \infty} \mu_k = \mu^*,$$

we see that Theorem 3 is proven with  $\widehat{u} = U^* - \xi$  and  $\lambda = \mu^*$ .

We mention that the same method allows the extension of the theorem by Kolmogorov and Arnold on Hamiltonian systems to the differentiable case. The perturbation has to be small in the  $\mathcal{C}^l$  topology where

$$l > 2 + 2(\tau + 1) > 2n + 2.$$

## REFERENCES

1. J. SCHAUDER, *Das Anfangswertproblem einer quasi-linearen hyperbolischen Differentialgleichung zweiter Order in beliebiger Anzahl von unabhängigen Veränderlichen*. Fund. Math. **24**, 1935, pp. 213-246.
2. K. O. FRIEDRICHS, *Symmetric Hyperbolic Linear Differential Equations*. Comm. Pure Appl. Math. **7**, 1954, pp. 345-392.
3. I. NASH, *The Imbedding of Riemannian Manifolds*, Ann. Math. **63**, 1956, pp. 20-63.
4. J. SCHWARTZ, *On Nash's Implicit Function Theorem*, Comm. Pure Appl. Math. **13**, 1960, pp. 509-530.
- 4'. J. MOSER, *A New Technique for the Construction of Solutions of Nonlinear Differential Equations*. Proc. Nat. Acad. Sciences, Vol. **47**, No. 11, pp. 1824-1831, 1961.
- 4''. J. MOSER, *On Invariant Curves of Area-Preserving Mappings of an Annulus*. Nachr. Akad. Wiss. Göttingen Math. Phys. K1. IIa, No. 1, 1962, pp. 1-20.
5. G. L. SIEGEL, *Iteration of Analytic Functions*. Ann. of Math. **43**, 1942, pp. 647-612.
- 5'. C. L. SIEGEL, *Über die Normalform Analytischer Differentialgleichungen in der Nähe einer Gleichgewichtslösung*, Nachr. Akad. Wiss., Göttingen, Math. Phys. K1, IIa, 1952, pp. 21-30.
6. A. N. KOLMOGOROV, Doklady Akad., Nauk USSR **98**, 1954, pp. 527-530.
7. A. N. KOLMOGOROV, *General Theory of Dynamical Systems and Classical Mechanics*. Proc. of Intl. Congress of Math., Amsterdam, 1954, (Amsterdam: Erven P. Noordhoff, 1957), Vol. 1, pp. 315-333.
8. V. I. ARNOLD, « *Small divisors* » I, *On mappings of a circle onto itself*, Izvest. Akad. Nauk, ser. mat., **25**, No. 1, 1961, pp. 21-76.
9. V. I. ARNOLD, *Small Divisor and Stability Problems in Classical and Celestial Mechanics*. Uspekhi Mat. Nauk USSR, Vol. 18, Ser. 6 (119), 1963, pp. 81-192.
10. V. I. ARNOLD, *Proof of A. N. Kolmogorov's Theorem on the Preservation of Quasi-Periodic Motions under Small Perturbation of the Hamiltonian*. Uspekhi Mat. N. USSR, Vol. 18, Ser. 5 (113), 1963, pp. 13-40.
11. K. O. FRIEDRICHS, *Symmetric Positive Linear Differential Equations*, Comm. Pure Appl. Math. **11**, 1958, pp. 333-418.
12. N. N. BOGOLIOUBOV, and Y. A. MITROPOLSKI, *The Method of Integral Manifolds in Non-linear Mechanics*, Contrib. to Differential Equations, Vol. II, 1963, pp. 123-196.
13. S. P. DILIBERTO, *Perturbation Theorems for Periodic Surfaces I*, Rend. del Circ. Mat. Palermo (2), **9**, 1961, pp. 265-299. II, Rend. del Circ. Mat. Palermo (2), **10**, 1962, pp. 111-161.
- 13'. S. P. DILIBERTO, *Perturbation Theory of Invariant Surfaces I-IV*. Mimeographed ONR Reports, Berkeley, 1956/57.
14. KYNER, W. T., *Invariant manifolds*, Rend. del Circolo Matematico Palermo, Ser. II, Vol. **9**, 1961, pp. 98-110.
- 14'. J. K. HALE, *Integral Manifolds of Perturbed Differential Systems*, Annals of Math., **73**, 1961, pp. 496-531.
15. I. KUPKA, *Stabilité des variétés invariantes d'un champ de vecteurs pour les petites perturbations*. Compt. Rend. Acad. Sc. Paris, **258**, Group 1, 1964, pp. 4197-4200.

16. L. NIRENBERG, *On elliptic partial differential equations*. Ann. Scuola Norm. Sup. Pisa, Ser. 3, **13** (1959), pp. 116-162.
17. GAGLIARDO, E., *Ulteriori proprietà di alcune classi di funzioni in più variabili*, Ricerche Mat, **8**, 1959, pp. 24-51.
18. J. J. KOHN and L. NIRENBERG, *Non-coercive boundary value problems*, to appear in Comm. Pure Appl. Math. (See under « elliptic regularization »).
19. W. T. KYNER, *A fixed point theorem*. Contributions to the Theory on Nonlinear Oscillations, Vol. 3, Princeton University Press, Annals of Math. Studies 36, 1956, pp. 197-205.
20. W. T. KYNER, *Small Periodic Perturbations of an Autonomous System of Vector Equations*. Contributions to the Theory of Nonlinear Oscillations, Vol. 4, Princeton, 1958.
21. R. SACKER, *On Invariant Surfaces and Bifurcation of Periodic Solutions of Ordinary Differential Equations*, NYU Report, IMM-NYU 333, Oct. 1964, to be published in Comm. Pure Appl. Math.
22. C. L. SIEGEL, *Vorlesungen über Himmelsmechanik*, Springer, 1956.
23. H. CREMER, *Über die Häufigkeit der Nichtzentren*, Math. Ann. **115**, pp. 607-612, 1942.
24. N. LEVINSON, *Transformation of an analytic function of several variables to a canonical form*. Duke Math. Journal **28**, 1961, pp. 345-353. (See Theorem 2).
25. N. I. ACHIESEER, *Vorlesungen über Approximationstheorie*, Akd. Verlag. Berlin 1953, translated from Russian, Chapter V.