M P E J

Mathematical Physics Electronic Journal

ISSN 1086-6655
Volume 2, 1996

Paper 4
Received: June 19, 1996, Accepted: June 21, 1996
Editor: G. Gallavotti

ABSOLUTELY CONVERGENT SERIES EXPANSIONS FOR QUASI PERIODIC MOTIONS.

L. H. Eliasson

Dept. of Math., KTH, Stockholm

CONTENTS.

I. INTRODUCTION.

The problem of finding formal quasi-periodic series in a small parameter ε of the solutions of certain analytic differential equations, depending on ε, occupied many astronomers and mathematicians during the last century (Lindstedt, Gyldén, Poincaré). What they were looking for was formal series in ε , which have coefficients that are analytic quasi-periodic functions, and which satisfy the differential equation *as if the series were convergent.* One version of this problem is to find quasi-periodic series for a fixed frequency ω, in which case one requires that all the coefficients are quasi-periodic with frequency ω.

In general, such series exist if one supplies a certain number of constants to the equation, and the problem then becomes one of *"killing the constants"* for the particular equation or class of equations considered. This problem was solved by Poincaré who showed that, for a generic class of Hamiltonian systems, there exists a formal solution for any fixed frequency ω for which such solutions exist when $\varepsilon = 0$. (Though, for example, the equations of the planetary problem in celestial mechanics are not generic in this sense, Poincaré's work applies also to them, and he established the existence of a formal solution also in this case. The frequencies are not, however, determined by the 0:th order approximation - the Kepler approximation - but by the first order approximation - the Lagrange approximation.)

Poincaré constructed in several ways *quasi-periodic series expansions of the formal solution*, i.e. convergent trigonometric series expansions for each of the coefficient of the formal solution, which he named after Lindstedt who made a major contribution to the solution of the problem. It is important to note that a quasi-periodic series expansion of the formal solution for a fixed frequency is not unique. This in contrast to the formal solution itself, for which we shall reserve the term *Lindstedt series.*

We shall now describe the problem more formally.

The Hamiltonian problem.

We first formulate the generic condition on the frequency vector ω in $\mathbf{R}^\nu$. This is the *Diophantine condition*

$$| \langle w, \omega \rangle | \geq \frac{1}{K \mid w \mid^\tau}, \quad w \in \mathbf{Z}^\nu \setminus 0$$

for some $\tau > \nu - 1$ and some positive constant K. The norm is $\mid w \mid = \mid w^1 \mid + ... + \mid w^\nu \mid$.

Let

$$h_0(x,y) = \langle \omega, y \rangle + \frac{1}{2} \langle y, Q(x) y \rangle + \mathbf{O}^3(y), \quad (x,y) \in \mathbf{T}^\nu \times \Omega_1 \subset \mathbf{T}^\nu \times \mathbf{R}^\nu$$

where Q is symmetric, and consider the perturbed analytic Hamiltonian $h_0(x,y) + \varepsilon^2 h_1(x,y,\varepsilon)$. By scaling the variables y by ε and then dividing the Hamiltonian by ε, we get a function of the form $\langle \omega, y \rangle + \varepsilon h(x,y)$, where h depends on ε and is equal to $h_1(x,0,0) + \frac{1}{2}\langle y, Q(x) y \rangle$ for $\varepsilon = 0$.

If we let $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$, then $t \to (x(t), y(t)) = (t\omega, 0) + X(t\omega)$ is a quasi periodic solution to the Hamiltonian system

$$\frac{dx}{dt} = \omega + \varepsilon \frac{\partial h}{\partial y}, \qquad \frac{dy}{dt} = -\varepsilon \frac{\partial h}{\partial x}$$

if and only if X satisfies

$$\partial X(x) = \varepsilon J h'((x,0) + X(x))$$

where ∂ is the directional derivative in the direction ω. It is far from obvious that this equation has a formal solution in ε - a Lindstedt series - but, as we have remarked above, this is indeed the case in general. The generic condition for this is that $det < Q > \neq 0$, where $< Q >$ is the mean value over $\mathbf{T}^\nu$.

Once the formal problem settled, Poincaré raised the question of convergence, i.e. if the Lindstedt series corresponds to really existing quasi-periodic solutions or not. He showed that the general formal quasi-periodic solution with varying frequencies (which he also showed to exist) is divergent, or, more precisely, not uniformly convergent in initial conditions and ε. The divergence being an effect of the well-known *small* *divisors.* The question remained, however, if the Lindstedt series for a fixed frequency (corresponding to fixed initial conditions) is convergent, uniformly with respect to ε. In "Les méthodes nouvelles de la mécanique célèste", vol II, section 149, [1] he discusses this question in detail without being able to decide it.

An essential part of the difficulty in proving the convergence of the Lindstedt series resides in the fact that it is given to us by a quasi-periodic series expansion which most naturally turns out to be *absolutely divergent* (in a sense we shall explain). So if the Lindstedt series converges, there must be very sharp *compensations of signs* between different terms of these series expansions. But since the absolute divergence is so fast, these compensations must be extremely precise. This led Poincaré to the conclusion that convergence is extremely unlikely. Today we know, through the works of Kolmogorov, Arnold and Moser [2, 3, 4] in the fifties and early sixties, that, for a fixed generic frequency, the Lindstedt series is convergent. (The so-called KAM-technique used for the proof of this consists of an iteration process on function spaces and has no relation to formal series at all.)

In this paper we shall describe the Lindstedt series together with a natural quasi-periodic series expansion, and we shall explain why it is absolutely divergent. We shall then describe a large class of compensations for the terms in the series expansion, and by taking into account these compensations, we shall obtain new quasi-periodic series which are absolutely convergent. This we shall prove by generalizing *Siegel's method* [5], a method developed by Siegel in the early forties which gives very good estimates of certain products of small divisors. In a second step we shall prove that these new series are, indeed, an expansion of the formal solution, which thus must be convergent. In order to do this, we shall resurrect the constants. As we mentioned above, supplying constants makes formal quasi-periodic expansions possible (for Hamiltonian systems and others). We shall show that, supplying even more constants makes these expansions (for a fixed generic frequency) absolutely convergent. Only then, the problem of "killing the constants" comes in.

Content of the paper.

In section II we shall describe the Lindstedt series and discuss its convergence properties.

In section III we describe Siegel's method. It consists of 3 lemmas. The first lemma gives an estimate of certain - *linear* - products of small divisors, and the third lemma gives an estimate of more general - *non-linear* - products. The step between these two results uses an arithmetical property which is described in his second lemma.

The products estimated by Siegel's method are of a very particular type, which we may call *non-resonant.* In the expansion of the Lindstedt series, however, there appears also *resonant* products for which Siegel's estimates do not hold. In section IV we shall therefore undertake a substantial generalization of Siegel's first lemma in order to get good estimates not of resonant linear products of small divisors but of certain sums of such products (proposition 4). In section V we apply Siegel's idea in order to get a generalization also of his third lemma, thereby obtaining good estimates for certain sums of resonant non-linear products (proposition 5).

We have not tried to formulate these estimates in the most general way but rather aimed at a formulation which is relevant for the Hamiltonian problem. Others are of course possible. In [6], for example, another generalization, which is relevant for linearization of a vector field on a torus, is given but not proven. This will follow as a special case of the estimates we give in this paper.

In section VI we describe new series which are absolutely convergent. Then we show that these new series also are an expansion of the formal solution which, thus, is shown to be convergent.

Acknowledgment. This work was done during several visits at the IHES, Bures-sur-Yvette, for whose hospitality the author is grateful.

II. THE CLASSICAL APPROACH TO THE HAMILTONIAN PROBLEM - ABSOLUTELY DIVERGENT SERIES.

Let $F = F(x, y)$ be an analytic mapping $\mathbf{T}^\nu \times \Omega \to \mathbf{R}^\mu$, where $\mathbf{T} = \mathbf{R}/(2\pi\mathbf{Z})$ and Ω is a neighborhood of the origin in $\mathbf{R}^\mu$. F may depend analytically on certain parameters. Let $\partial = \langle \nabla \ , \omega \rangle$ be the derivative in the direction ω, and consider the equation

$$(*) \qquad \partial X(x) = \varepsilon F(x, X(x)) - C, \quad < X >= 0$$

for a mapping $X : \mathbf{T}^\nu \to \mathbf{R}^\mu$ and a constant vector C in $\mathbf{R}^\mu$. X and C will, of course, depend on the parameter ε which is assumed to be small.

Supplying the constant C is necessary in order to have a solution X since the left hand side has mean value 0, and, therefore, the right hand side also must have a mean value that vanishes. Fixing the mean value $< X >= \frac{1}{2\pi}\int_{\mathbf{T}^\nu} X(x)dx$ is necessary in order to have uniqueness, and, since we admit that F can depend on parameters, it is no restriction to let it be 0.

The problem of finding quasi-periodic solutions to a Hamiltonian system can be formulated as equation (*) when $\mu = 2\nu$ and

$$F(x, y) \;=\; Jh'((x, 0) + y)$$

where $h(x, x')$ is defined on $\mathbf{T}^\nu \times \Omega_1$, Ω_1 open in $\mathbf{R}^\nu$, and Jh' is the symplectic gradient of h, i.e. the Hamiltonian vector field.

The formal solution and its series expansion.

There exists a *formal solution* of (*)

$$X(x) \;\sim\; \sum_{k\geq 1} \varepsilon^k X_k(x) \tag{1$_1$}$$

$$C \;\sim\; \sum_{k\geq 1} \varepsilon^k C_k \tag{1$_2$}$$

i.e. power series in ε which, when evaluated formally in (*), give an identity for any power of ε. This formal solution is unique, i.e. the functions $X_k(x)$ and the vectors C_k are uniquely determined. What interest us here is to estimate them. For this we shall study *the series expansion of $X_k(x)$ and C_k in terms of the given data $F(x, y)$*. This series expansion is obtained by trivial comparison of coefficients in the equation $(*)$, but to describe it we must introduce some notations.

Let $F(x, y) = \sum_{j\geq 0} F_j(x)(y)^j$ and let $\underline{\hat{F}(j, w)} = \hat{F}_j(w)$, $w \in \mathbf{Z}^\nu$, be the w:th Fourier coefficient of F_j. So $\hat{F}(j, w)$ is a (symmetric) j-linear mapping. Then

$$X_k(x) \;=\; (\sqrt{-1})^{-k} \sum \Lambda_1(\delta, v)\hat{F}(\delta, v) e^{\sqrt{-1}\langle v_1+\ldots+v_k, x\rangle} \tag{2$_1$}$$

$$C_k \;=\; (\sqrt{-1})^{-k+1} \sum \Lambda_2(\delta, v)\hat{F}(\delta, v) \tag{2$_2$}$$

where summation runs over all $v = (v_1, \ldots, v_k) \in \underline{\Gamma(k)} = (\mathbf{Z}^\nu)^k$ and $\delta = (\delta_1, \ldots, \delta_k) \in \underline{\Delta(k)} \subset \mathbf{N}^k$.

In order to give meaning to this expression, we define the composition of two multilinear maps L_j and L_i, $j \geq 1$, by

$$L_j L_i(X_1, \ldots, X_{i+j-1}) = L_j(L_i(X_1, \ldots, X_i), X_{i+1}, \ldots, X_{i+j-1})$$

and we let $L_k L_j L_i = (L_k L_j)L_i$ (which is equal to $L_k(L_j L_i)$ whenever this is defined, i.e. for $j, k \geq 1$). So

$$\underline{\hat{F}(\delta, v)} \;=\; \hat{F}(\delta_k, v_k) \ldots \hat{F}(\delta_1, v_1)$$

is a well-defined multi-linear mapping, and is also a vector, if

$$\sum_{j\le i\le k} \delta_i \;\ge\; k-j+1 \;\; (1<j\le k) \qquad and \qquad \sum_{1\le i\le k} \delta_i \;=\; k-1.$$

$\Delta(k)$ is the set of all sequences $(\delta_1,...,\delta_k)$ for which these k conditions hold.

If we now introduce these series in (*) and identify the coefficients of the formal "monomials" $\hat{F}(\delta_k,v_k)....\hat{F}(\delta_1,v_1)$, we obtain recurrence relations for $\Lambda_1(\delta,v)$ and $\Lambda_2(\delta,v)$ which determine these numbers uniquely. So there is a unique set of numbers $\Lambda_1(\delta,v)$ and $\Lambda_2(\delta,v)$ such that (1)+(2) solves (*) formally for any F.

But the series (1) are unique also in a stronger sense. In fact, the X_k:s are well-defined functions and the C_k:s are well-defined vectors (since ω is Diophantine), and they are uniquely determined by F. For a fixed F, however, the particular series expansion (2) of X_k and C_k is not unique, i.e the coefficients $\Lambda_1(\delta,v)$ and $\Lambda_2(\delta,v)$ are not uniquely determined by F. (They are independent of F!) For example, if F is linear in y with $F_1(x)$ nilpotent, then there are many other series of the form (2) which represent the same formal solution.

"Killing the constants" and the Lindstedt series.

In the Hamiltonian case, (1)+(2) gives a formal solution to

$$\partial X(x) = \varepsilon J h'((x,0)+X(x)) - C, \quad <X>=0.$$

C, however, is not 0 in general. But one can introduce 2ν parameters $<X>=(\lambda_1,\lambda_2)$ with which one may hope to kill the vector $C\in \mathbf{R}^{2\nu}$. It is easy to see, however, that C is independent of λ_1 and that we therefore only have ν parameters available. So "killing the constants" is here a real problem. That it can be solved, under the condition $det<Q>\neq 0$, was shown by Poincaré by an argument which uses the symplectic character of the problem. We shall now describe this argument.

We have a formal solution (1)+(2) to

$$\partial X(x) = \varepsilon J h'((x,\lambda_2)+X(x)) - C, \quad <X>=0 \tag{3}$$

where each X_k and C_k depend analytically on λ_2. We must show that we can determine λ_2 as a formal series in ε in such a way that C vanishes.

If we write

$$C = \begin{pmatrix} C^1 \\ C^2 \end{pmatrix}$$

and if we observe that

$$C = \varepsilon \begin{pmatrix} <Q>\lambda_2 \\ 0 \end{pmatrix}$$

modulo ε^2, we can solve

$$C^1 = 0 \tag{4}$$

formally and uniquely for $\lambda_2=\lambda_2(\varepsilon)$, since $det<Q>\neq 0$. A priori, this kills only a part of the constants, and it may seem likely that not all C vanish for λ_2 so determined. But, indeed it does.

Lemma 1. $C=0$ for $\lambda_2=\lambda_2(\varepsilon)$.

Proof. In order to show this, we must use the symplectic character of the problem. From classical perturbation theory we have the following result due to Poincaré ([1], section 126):

> There is a formal mapping $\Phi(x,y)=\sum_{k\ge1}\varepsilon^k\Phi_k(x,y)$, each Φ_k being analytic in $x\in\mathbf{T}^\nu$ and in y, λ_2 for small values of y and λ_2, and there is a formal series $f(y)=\sum_{k\ge0}\varepsilon^k f_k(y)$, each f_k being analytic in y and λ_2, such that $\tilde{\Phi}(x,y)=(\tilde{\Phi}^1(x,y),\tilde{\Phi}^2(x,y))=(x,y)+\Phi(x,y)$ is exact symplectic, i.e.
>
> $$\tilde{\Phi}^*(\sum y_i dx_i) = \sum y_i dx_i$$

and

$$\langle\omega,\tilde{\Phi}^2(x,y)\rangle+\varepsilon h(\tilde{\Phi}(x,y)+(0,\lambda_2))=\langle\omega,y\rangle+\varepsilon f(y).$$

The series $f(y)$ is unique. $\Phi(x,y)$ is not unique, but gets uniquely determined by the additional condition that $<\Phi^1(\ ,y)>=0$ for all y.

Differentiating the above equality once, using that $\tilde{\Phi}$ is symplectic and letting $y=0$, we get, for $\tilde{X}(x)=\Phi(x,0)-\begin{pmatrix}0\\<\Phi^2(\ ,0)>\end{pmatrix}$,

$$\partial\tilde{X}(x)=\varepsilon Jh'((x,\lambda_2+<\Phi^2(\ ,0)>)+\tilde{X}(x))-\varepsilon Jf'(0)-\varepsilon\Phi'(x,0)Jf'(0),\quad <\tilde{X}>=0$$

where $f'(y)$ is the gradient of f considered as a function of x and y, i.e. $f'(y)=\begin{pmatrix}0\\\frac{\partial f}{\partial y}(y)\end{pmatrix}$.

Now $\frac{\partial f}{\partial y}(0)=<Q>\lambda_2+\mathbf{O}(\varepsilon)$, and, since $<Q>$ is non-singular, we can define a unique formal series $\tilde{\lambda}_2(\varepsilon)$ such that $f'(0)=0$ for $\lambda_2=\tilde{\lambda}_2(\varepsilon)$.

With λ_2 so chosen, we get a formal solution $\tilde{X},\ \tilde{C}=0$ to the equation

$$\partial\tilde{X}(x)=\varepsilon Jh'((x,\tilde{\lambda}_2(\varepsilon)+<\Phi^2(\ ,0)>)+\tilde{X}(x))-\tilde{C},\quad <\tilde{X}>=0.$$

This formal equation has a unique formal solution. But X,C, for $\lambda_2=\tilde{\lambda}_2(\varepsilon)$, is also a formal solution of this equation, so we conclude that

$$X=\tilde{X},\ C=\tilde{C}.$$

Hence $C=0$ for $\lambda_2=\tilde{\lambda}_2(\varepsilon)+<\Phi^2(\ ,0)>$.

This implies that the formal series $\tilde{\lambda}_2(\varepsilon)+<\Phi^2(\ ,0)>$ solves the equation (4). By the uniqueness of the solution of (4), it follows that

$$\tilde{\lambda}_2(\varepsilon)=\lambda_2(\varepsilon)-<\Phi^2(\ ,0)>.$$

This proves the lemma. ∎

Hence, the Hamiltonian problem has a formal solution $(1)_1$ when the mean value λ_2 is determined according to (4) - the Lindstedt series. If we now ask about the convergence of the Lindstedt series, the most natural thing to do is to consider the explicit expansion given by $(2)_1$. In order too describe this expansion we shall first introduce some notations.

Index sets.

Definition. A *simple index set* is a finite subset A of $\mathbf{Z}$ together with a $\delta\in\Delta(k)$, $k=\#A$. An *index set* is a disjoint union of simple index sets.

Notice that a subset of $\mathbf{Z}$ is canonically isomorphic to $\{1,...,k\}$ for some k, so we can always assume that A is such a set. Or, equivalently, we can consider δ as a mapping $A\to\mathbf{N}$.

Lemma 2. For any simple index set A,δ there is a unique (partial) ordering $\prec$ on A such that

(i) $\delta(x)=\#\{immediate\ predecessors\ of\ x\ in\ A,\prec\}$ (y is a *predecessor* of x if $y\prec x$, and it is *immediate* if for no z it holds that $y\prec z\prec x$);

(ii) $y\prec x$ implies $y<x$;

(iii) $y\prec x$ and $y<z<x$ implies $z\prec x$.

Conversely, any ordering $\prec$ satisfying (ii) and (iii) is of this form.

Proof. The proof is an easy induction on $\#A=k$.

Notice that $\delta(1)=0$ and let x be the smallest element (with respect to $<$) in A such that $r=\delta(x)\geq 1$. Then $x-r,x-(r-1),...,x-1$ are the immediate predecessors of x. Let $A'=A\setminus\{x-r,...,x-1\}$ and define $\delta_{A'}\in\Delta(k-r)$ by $\delta_{A'}(y)=\delta(y)$ if $y\in A'\setminus\{x\}$ and $\delta_{A'}(x)=0$. The first part now follows by induction.

To prove the second part, let $\prec$ be an ordering satisfying the assumptions. Let $x\in A$ have immediate, but no other, predecessors (such elements always exists if $\#A\geq 1$), and let $B=\{x\}\cup\{predecessors\ of\ x\}$

and $A' = A \setminus \{predecessors\ of\ x\}$. By induction, there are $\delta_{A'} \in \Delta(k-r)$ and $\delta_B \in \Delta(r+1)$, $r+1 = \#B$. We now define $\delta \in \Delta(k)$ by $\delta(y) = \delta_{A'}(y)$ if $y \in A' \setminus \{x\}$, $\delta(y) = \delta_B(y)$ if $y \in B \setminus \{x\}$, and $\delta(x) = r$. ∎

We say that A, δ is a *linear index set* if $\prec$ is a total ordering, i.e. if $\delta(1) = 0,\ \delta(2) = ... = \delta(k) = 1$.

A *sub index set* of A, δ_A is a subset $B \subset A$ together with $\delta_B \in \Delta(r)$, $r = \#B$, which is defined by the induced ordering on B according to lemma 2. Notice that $\delta_B \neq \delta_A/B$ in general.

Any subset of A is again an index set for the induced ordering. In particular, for any $x, y \in A,\ x \preceq y$, the sets

$$\underline{[x,y]} = \{z \in A : x \preceq z \preceq y\}, \quad \underline{A(x)} = \{z \in A : z \preceq x\}$$

are sub index sets. Other such sets are $]x,y]$, $[x,y[$ and $]x,y[$ which are equal to $[x,y]$ minus $\{x\}$,$\{y\}$ and $\{x,y\}$, respectively. Notice that the sets $[x,y]$, $]x,y]$ etc. are totally ordered for the induced ordering.

For x in an index set A we let $\underline{'x}$ denote the unique immediate predecessor of x if such an element exists. The immediate successor of x is always unique, if it exists, and will be denoted by $\underline{x'}$. Throughout this paper, *any primed indices will have this meaning.*

An index set has a *natural decomposition* into simple sub index sets:

$$A = A(b^s) \cup ... \cup A(b^1), \quad s \geq 2$$

if A is not simple ($b^1 < ... < b^s$ being the maximal elements);

$$A = \{b\} \cup A(b^s) \cup ... \cup A(b^1), \quad s \geq 0$$

if A is simple (b being the maximal element in A and $b^1 < ... < b^s$ being its immediate predecessors).

The number of different index sets on A is not very large.

Lemma 3. $\#\Delta(k) \ \leq\ 4^k$.

Proof. Notice that $\#\Delta(k) \leq N(k,k)$, where

$$N(k,n) \ = \ \#\{(x_1, ..., x_n) \in \mathbf{N}^n : x_1 + ... + x_n = k\}.$$

An easy induction gives that $N(k,n) \leq 2^{n+k}$. ∎

Description of the coefficients.

Let now $A = \{1, ..., k\}$ be provided with the ordering defined by δ. Then we associate to $v : A \to \mathbf{Z}^\nu$ a mapping $\gamma = \gamma_{\delta,v} : A \to \mathbf{C}$ through

$$\underline{\gamma_{\delta,v}(x)} \ = \ \langle \sum_{y \preceq x} v(y), \omega \rangle.$$

This representation is a bijection. So we can, given δ, represent v by γ in an unambiguous manner. Notice that this representation behaves well for restrictions to subsets B of the form $B = A(x)$, but not to arbitrary subsets. For example, if $A = [a,b]$ and $B =]c,d[$, and if v is represented by γ, then v/B is represented by $(\gamma - \gamma(c))/B$, while γ/B represents the mapping $\tilde{v}$, where $\tilde{v}(c') = \sum_{y \preceq c'} v(y)$ and $\tilde{v} = v$ otherwise.

Let's now return to the description (2) of the formal solution. Let $(\delta, v) \in \Delta(k) \times \Gamma(k)$, and consider the natural decomposition of $A = \{1, ..., k\}$ with respect to δ:

$$A \ = \ \{b\} \cup A^s \cup ... \cup A^1, \quad s \geq 0$$

where $b = k$ is the maximal element. Let $\gamma = \gamma_{\delta,v}$, and notice that $\gamma(b)$ vanishes if and only if $\sum_{x \in A} v(x) = 0$.

Clearly $\Lambda_1(\delta, v) = 0$ when $\gamma(b) = 0$, and $\Lambda_2(\delta, v) = 0$ when $\gamma(b) \neq 0$. Then equation (*) gives the following recurrence relations for the coefficients:

$$\gamma(b)\Lambda_1(\delta,v)\hat{F}(\delta,v) \ = \ \hat{F}(\delta(b), v(b))(V_s, ..., V_1), \qquad V_i = \Lambda_1((\delta,v)/A^i)\hat{F}((\delta,v)/A^i)$$

if $\gamma(b) \neq 0$;

$$0 \; = \; \hat{F}(\delta(b), v(b))(V_s, ..., V_1) \; - \; \Lambda_2(\delta, v)\hat{F}(\delta, v), \qquad V_i = \Lambda_1((\delta, v)/A^i)\hat{F}((\delta, v)/A^i)$$

if $\gamma(b) = 0$. ($(\delta, v)/A^i$ is of course to be understood as the restriction of the mapping $\delta, v : A \to \mathbf{N} \times \mathbf{Z}^\nu$ to the subset $A^i \subset A$.)

From this we easily get

$$\begin{aligned} \Lambda_1(\delta, v) &= \begin{cases} (\prod_{x \in A} \gamma(x))^{-1} & \text{if } \gamma(x) \neq 0 \text{ for all } x \in A \\ 0 & \text{otherwise} \end{cases} \\ \Lambda_2(\delta, v) &= \begin{cases} \Lambda_1(\delta, v/A \setminus \{b\}) & \text{if } \gamma(b) = 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{5}$$

where $\Lambda_1(\delta, v/A \setminus \{b\})$ is the product $\prod \Lambda_1(\delta, v/A^i)$.

Convergence and divergence of the formal solution.

The series (2) are absolutely convergent if ω is Diophantine, as we have assumed. By this we mean that

$$\sum_{\delta \in \Delta(k)} \sum_{v \in \Gamma(k)} \mid \Lambda_i(\delta, v) \mid\mid \hat{F}(\delta_k, v_k) \mid .... \mid \hat{F}(\delta_1, v_1) \mid \tag{6}$$

converges, $i = 1, 2$. In fact, there are numbers $0 < s, r < 1$ such that

$$\mid \hat{F}(j, w) \mid \leq C_F s^{-j} e^{-|w| r}$$

for all j and w, where the constant C_F depends on the supremum norm of F over some complex neighborhood of $\mathbf{T}^\nu \times 0$.

Since $\#\Delta(k) \leq 4^k$, the series (6) can be estimated by

$$(4C_F)^k s^{-k+1} \sum_{v \in \Gamma(k)} \sup_{\delta \in \Delta(k)} \mid \Lambda_i(\delta, v) \mid e^{-(|v_1| + ... + |v_k|) r}.$$

And this series converges if $\sup_{\delta \in \Delta(k)} \mid \Lambda_i(\delta, v) \mid$ is a polynomial in $\mid v_1 \mid + ... + \mid v_k \mid$, which is the case as we see from (5). Hence, the series (2) converge absolutely to X_k and C_k for each k.

By *absolute convergence* of (1)+(2) we mean convergence of (1) when $X_k(x)$ and C_k are replaced by (6). This is clearly sufficient (but not necessary) for the convergence of (1). But such absolute convergence does not take place in general. To see this, it suffices to consider the linear case.

Example. If $A = \{1, ..., k\}$ and δ is a linear index set, then

$$\Lambda_1(\delta, v) = (\langle v_1, \omega \rangle \langle v_1 + v_2, \omega \rangle ... \langle v_1 + v_2 + ... + v_k, \omega \rangle)^{-1}$$

whenever this is defined, and $= 0$ otherwise.

Let now w be such that $\mid \langle w, \omega \rangle \mid < C \mid w \mid^{-\nu + 1}$ and choose $v_1 = w$, $v_2 = v_4 = ... = v_k = e = (1, 0, ..., 0)$ and $v_3 = v_5 = ... = v_{k-1} = -e$ (assuming k even). Then

$$\mid \Lambda_1(\delta, v) \mid = \mid \langle w, \omega \rangle \mid^{-\frac{k}{2}} \mid \langle w + e, \omega \rangle \mid^{-\frac{k}{2}} \; > \; (C' \mid w \mid^{\frac{\nu - 1}{2}})^k.$$

If now $\mid w \mid = k$ (such w can always be found for arbitrary large k by a well known theorem of Dirichlet - see for example [7]), then we get

$$\mid \Lambda_1(\delta, v) \mid e^{-(|v_1| + ... + |v_k|) r} \; > \; (C'' \mid k \mid^{\frac{\nu - 1}{2}} e^{-2r})^k.$$

Hence, (6) increases much to fast to be compensated by a factor ε^k. This shows that there is no absolute convergence. (One would, of course, obtain an even worse estimate by letting $v_2 = ... = v_k = 0$. This case is less serious, however, since in many cases $\hat{F}_1(0)$ is 0 or nilpotent, as we shall see below.)

Notice that the very large value of $\Lambda_1(\delta, v)$, compared with $\varepsilon^k \mid \hat{F}(\delta, v) \mid$, has been obtained by repetition of many very bad small divisors in the product - a resonant product. If no such repetitions occur - a non-resonant product - then we have a much better estimate (lemma 4).

Remark. In order for (1) to converge for certain (classes of) F, $\mid X_k(x) \mid$ and $\mid C_k \mid$ must be much smaller than (6). This requires that there are *compensations of signs* between different terms in (2). But since (6) diverges so fast, these compensations must be very precise. This, however, is precisely what happens in the Hamiltonian case - there is conditional convergence of (1)+(2) with compensations of signs, but no absolute convergence.

III. SIEGEL'S METHOD.

Many of the coefficients $\Lambda_1(\delta, v)$ have a very good estimate. If $\gamma_{\delta,v}(x) \neq \gamma_{\delta,v}(y)$ whenever $x \prec y$, then we have the following result due to Siegel [5]:

$$\mid \Lambda_1(\delta, v) \mid \leq (2^{3\tau+1}K)^k \prod_{v_i \neq 0} \mid v_i \mid^{3\tau} . \tag{7}$$

In the series (1)+(2), however, there are many terms for which $\gamma_{\delta,v}$ does not have this property, and then this estimate cannot hold, as we have seen in the above example.

We shall now give Siegel's proof of (7).

Siegel's first lemma.

Lemma 4. Suppose that A, δ is linear, i.e. $\delta(1) = 0,\ \delta(2) = ... = \delta(k) = 1$. If $\gamma = \gamma_{\delta,v}$ is injective, then

$$\mid \Lambda_1(\delta, v) \mid \leq 2^{-\tau-1}(2^{\tau+1}K)^k \prod_{1 \leq j \leq k} \mid v(j) \mid^{\tau} .$$

Proof. We can assume that $\gamma \neq 0$, since otherwise $\Lambda_1(\delta, v) = 0$.

The estimate is obvious for $k = 1$, so we proceed by induction on k. Let x be such that

$$\mid \gamma(x) \mid = max\{\mid \gamma(j) \mid : j \in A\}.$$

Then

$$\mid \gamma(x) \mid \geq \frac{1}{2} max(\mid \gamma(x) - \gamma('x) \mid, \mid \gamma(x') - \gamma(x) \mid)$$

where, of course, these two inequalities reduce to one if $x = 1$ or k. By induction it follows that

$$\prod_j \mid \gamma(j) \mid^{-1} = \mid \gamma(x) \mid^{-1} \prod_{j \neq x} \mid \gamma(j) \mid^{-1} \leq 2^{-\tau-1}(2^{\tau+1}K)^k \prod_{1 \leq j \leq k} \mid v(j) \mid^{\tau} \times I$$

where

$$I = 2^{-\tau-1}K^{-1} \mid \gamma(x) \mid^{-1} (\frac{\mid v(x) + v(x') \mid}{\mid v(x) \mid\mid v(x') \mid})^{\tau}.$$

Since I is less than 1, the result follows. (This holds if $x \neq k$. When $x = k$ then the induction is even easier.) ∎

Remark. As we have seen above, this estimate is not valid without the assumption that γ is injective. However, one can weaken this assumption in the following way:

if, for any $i < j$ such that $\gamma(i) = \gamma(j)$, there is an $i < l < j$ such that $\mid \gamma(l) \mid \leq \mid \gamma(i) \mid$, then the estimate remains valid. (This is an easy variant of lemma 4.)

One can weaken the assumption even more to the cost of getting a bigger constant. For example,

if, for any $i < j$ such that $\gamma(i) = \gamma(j)$, there is an $i < l < j$ such that $\mid \gamma(l) \mid \leq 3 \mid \gamma(i) \mid$, then the estimate remains valid with the constant $2^{\tau+2}K$ instead of $2^{\tau+1}K$.

These simple observations will play an important role later.

Siegel's second lemma.

Lemma 5. Let $v_1, ..., v_r,\ r \geq 0$, and $u_1, ..., u_s,\ s \geq 2$, be positive integers such that

$$v_1 + ... + v_r + u_1 + ... + u_s = N,\ u_1 + ... + u_s \geq \frac{N}{2}\ and\ u_i \leq \frac{N}{2}.$$

Then

$$(v_1 ... v_r u_1^2 ... u_s^2)^{-1} \leq 2^r 4^s N^{-3}.$$

Proof. We shall prove the lemma under the weaker assumption that the v_i:s and u_i:s are real numbers larger than 1.

Since $v_1 + ... + v_r \leq 2^{r-1} v_1 ... v_r$, it suffices to prove the inequality for $r \leq 1$.

Since $(u_1 + u_2)^2 \leq 4(u_1 u_2)^2$, we can apply the same argument to $u_1 + u_2$ if $u_1 + u_2 \leq \frac{N}{2}$. Hence, we can assume that $u_i + u_j > \frac{N}{2}$ for all $i \neq j$. In particular, $s \leq 3$.

Suppose $s = 3$ and suppose $u_1 \leq u_2 \leq u_3$. Then u_2 is strictly less than $\frac{N}{2}$ and $(u_1 - t)(u_2 + t) \leq u_1 u_2$ for positive t. Hence, we can assume that $u_1 = 1$. By the same argument we can assume that $u_2 = 1$ or $u_3 = \frac{N}{2}$. Then $u_1 + u_2 \leq \frac{N}{2}$ (except when $N = 3$, in which case the result is trivial), so we can use the preceding argument to reduce this to the case $s = 2$.

If $s = 2$, then the above argument works to show that $u_1 = 1$ or $u_2 = \frac{N}{2}$.

For these different cases the verification of the lemma is immediate. ∎

Siegel's third lemma.

Lemma 6. Let $\gamma = \gamma_{\delta,v}$. If $\gamma(x) \neq \gamma(y)$ whenever $x \prec y$, then

$$\mid \Lambda_1(\delta, v) \mid \leq\ 2^{-\tau-1}(2^{3\tau+1}K)^k \prod_{\substack{1 \leq j \leq k \\ v_j \neq 0}} \mid v(j) \mid^{3\tau} (2 \sum_{1 \leq j \leq k} \mid v(j) \mid)^{-2\tau}.$$

Proof. We can assume that $\gamma \neq 0$, since otherwise $\Lambda_1(\delta, v) = 0$.

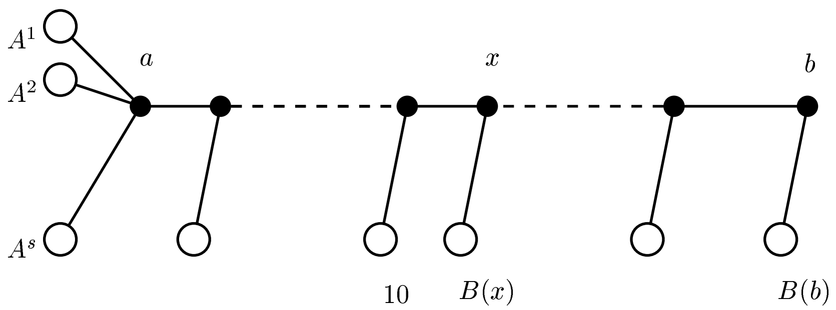
The lemma is true if $\#A = 1$, so we can proceed by induction on $\#A$.

Let

$$\overline{B} = \sum_{y \in B} v(y)\ and\ \mid B \mid = \sum_{y \in B} \mid v(y) \mid$$

for any subset B of A.

Consider the set $\{x : \mid A(x) \mid > \frac{1}{2} \mid A \mid\}$. This set is totally ordered, hence of the form $[a, b]$, where $A = A(b)$. We can decompose A according to $[a, b]$:



Here $A(x) = \{x\} \cup A('x) \cup B(x)$ for each $x \in]a,b]$, where $'x$ is the predecessor of x in $[a,b]$, and $A(a) = \{a\} \cup A^s \cup ... \cup A^1$, $s \geq 0$, is the natural decomposition. Each A^i is a simple non-void index set, while the $B(x)$:s may be void, simple or non-simple.

We shall use the following convention:

$\prod_I c_i$ is the product of all c_i, $i \in I$, which are $\neq 0$, and it is 1 if no such c_i exist. In particular, it is 1 if $I = \emptyset$. In agreement with this, we let $\prod c$ be c if $c \neq 0$, and be 1 if $c = 0$.

Now we have

$$\prod_{x \in A} \gamma(x) = (\prod_{a \leq x \leq b} \gamma(x)) \prod_B (\prod_{y \in B} \gamma(y))$$

where B runs over all non-void $B(x)$, $x \in]a,b]$ and all A^i. We can now apply lemma 4 to the first factor, and lemma 6 (by induction) to the others:

$$| \prod_{a \leq x \leq b} \gamma(x) |^{-1} \leq 2^{-\tau-1}(2^{\tau+1}K)^{\#[a,b]} \prod_{a < x \leq b} | v(x) + \overline{B(x)} |^\tau \times | \overline{A(a)} |^\tau$$

$$| \prod_{y \in B} \gamma(y) |^{-1} \leq 2^{-\tau-1}(2^{3\tau+1}K)^{\#B} \prod_{y \in B} | v(y) |^{3\tau} (2 | B |)^{-2\tau}.$$

Putting this together gives

$$2^{-\tau-1}(2^{3\tau+1}K)^{\#A} \prod_{x \in A} | v(x) |^{3\tau} (2 | A |)^{-2\tau} \times I_1 \times I_2$$

where

$$I_1 = (2^{-2\tau})^{\#[a,b]} (\frac{| \overline{A(a)} |}{\prod | v(a) |})^\tau (2 | A |)^{2\tau}$$

$$I_2 = (\prod_{a < x \leq b} (4 | B(x) |) \prod_{a < x \leq b} | v(x) | \prod_{1 \leq x \leq s} (8 | A^i |^2) \prod | v(a) |^2)^{-\tau}.$$

If $| v(a) | = 0$, then $s \geq 2$, and if $| v(a) | < \frac{1}{2} | A |$, then $s \geq 1$. In both cases we can use lemma 5 to estimate I_2. Hence,

$$I_1 \times I_2 \leq (2^{-2\tau})^{\#[a,b]} | A |^\tau (2 | A |)^{2\tau} (2^{\#]a,b]} \frac{4^{s+1}}{8^s} | A |^{-3})^\tau$$

which is less than 1. (Notice that $s \geq 2$ if $\#[a,b] = 1$.)

Suppose now $| v(a) | \geq \frac{1}{2} | A |$. If $s = 0$, then $| \overline{A(a)} | = | v(a) |$. If, moreover, all $B(x) = \emptyset$, $x \in]a,b]$, then we are in the linear case and the result follows from lemma 4. So we can assume that some $B(x)$ is non void and estimate I_2 by $2^{-2\tau} | v(a) |^{-2\tau}$. Then $I_1 \times I_2 \leq 1$.

If $s \geq 1$, then we can also estimate I_2 by $2^{-3\tau} | v(a) |^{-3\tau}$, and, hence, $I_1 \times I_2 \leq 1$.

This finishes the proof of the lemma. ∎

Remark. The constant $2^{3\tau+1}K$ is certainly not optimal and can easily be ameliorated. The exponent 3τ can be decreased to $\frac{7}{3}\tau$, but we don't know if this is the best.

Remark. Siegel's result gives absolute convergence of (1)+(2) in one particular case. This is when F is of "holomorphic type" and $\hat{F}_1(0) = 0$. By holomorphic type we mean that F is the restriction to $(\partial \mathbf{D})^\nu \times \Omega$ of a holomorphic function on $\mathbf{D}^\nu \times \Omega$, $\mathbf{D}$ being the unit disc in $\mathbf{C}$, i.e. $\hat{F}_j(w) = 0$ unless $w \in \mathbf{N}^\nu$. These assumptions imply that $\hat{F}(\delta, v) = 0$ whenever $\gamma_{\delta,v}$ does not fulfill the assumption of lemma 6. So we can assume that (7) holds for all $\Lambda_1(\delta, v)$, and we get

$$| X_k(x) | + | C_k | \leq [2^{3\tau+3} C_F K C_\tau \frac{1}{s r^{3\tau+\nu-1}}]^k$$

where C_τ is a constant that only depends on τ.

Hence, we have absolute convergence of (1)+(2). The assumptions on F, however, is very restrictive

Remark. Siegel's method has been applied to many problems of holomorphic type: linearization of holomorphic maps near an elliptic fixed point [5, 8, 9]; linearization of holomorphic vector fields near an elliptic stationary point [10]; construction of invariant symplectic submanifolds for a holomorphic symplectic mapping near an elliptic fixed point [11]. A variant of Siegel's method has been developed by Brjuno [12].

IV GENERALIZATION OF SIEGEL'S FIRST LEMMA.

Resonances on linear index set.

We assume that we are given an index set $A = A(b)$ together with a mapping $\gamma : A \to \mathbf{C}$. We assume that A is linear $= [a, b]$, though this will be of no importance.

Definition. A *γ-resonance* is a pair $(c, d) \in A \times A,\ c \prec d$, such that $\gamma(c) = \gamma(d)$ *and* $]c, d[\neq \emptyset$. A *short γ-resonance* is a pair $(c, d) \in A \times A,\ c \prec d$, such that $\gamma(c) = \gamma(d)$ *and* $]c, d[= \emptyset$.

According to this definition, a short γ-resonance is *not* a γ-resonance.

Let (e, f) be another γ-resonance. We say that (e, f) is *smaller than* (c, d) if $[e, f] \subset [c, d]$. "Smaller than" defines an ordering on the resonances, and notions like *maximal* or *largest* and *minimal* or *smallest* will refer to this ordering.

We say that (e, f) and (c, d) are *non-overlapping* if one is smaller than the other, or if they are *disjoint*. By this we mean that $]e, f] \cap]c, d] = \emptyset$. Notice that if (c, d) and (d, e) are resonances, then they are disjoint according to this definition.

A set or a *family of resonances* is called non-overlapping if all its elements are pairwise non-overlapping. A particular such family is *the trivial family* $J_\emptyset$ which consists of no resonances at all.

Let J be a family of non-overlapping family of resonances. We define its *support* to be

$$suppJ \ = \ \cup_{(c,d)\in J}]c, d[.$$

For any $x \in A$ we define $\underline{\gamma(J)(x)}$ in the following way:

if there is no $(c, d) \in J$ such that $x \in]c, d[$, let $\gamma(J)(x) = \gamma(x)$;

otherwise, take the smallest $(c, d) \in J$ such that $x \in]c, d[$, and let $\gamma(J)(x) = \gamma(x) - \gamma(c)$.

Example. As an illustration we consider $A = \{1, ..., 10\}$ with its natural ordering. Suppose $(1, 6)$, $(3, 5)$ and $(8, 10)$ are resonances with respect to γ, and let J consists of all these three resonances. Then the value of $\gamma(J)(x)$ is $\gamma(4) - \gamma(3)$ if $x = 4$, $\gamma(9) - \gamma(8)$ if $x = 9$, $\gamma(x) - \gamma(1)$ if $x = 2, 3, 5$, and $\gamma(x)$ for the other values of x.

Definition. J is said to be *admissible* if $\gamma(J)(x) \neq 0$ for all $x \in A$. We denote by $\underline{ad(\gamma)}$ the set of all admissible families.

Besides that γ may vanish, non-admissibility is due to the existence of "multiple" resonances. If, in the example above, $\gamma(3) = \gamma(1)$, then J is not admissible. But "multiple" resonances are not forbidden. For example $\gamma(4) = \gamma(1)$ does not prevent J from being admissible.

Remark. Some care must be taken with the concept of admissibility when we are considering "restrictions" to some subset B of A.

If J is a family of non-overlapping resonances in A (γ- resonances), let $\underline{J/B} = \{(c, d) \in J : c, d \in B\}$. Then clearly J/B is a family of non-overlapping resonances in B (i.e. γ/B-resonances). But J/B will in general not belong to $ad(\gamma/B)$ even if $J \in ad(\gamma)$. This is so since $\gamma(J/B) \neq \gamma(J)/B$ in general.

For "extensions", the situation is the same. If J is a family of non-overlapping resonances in B (i.e. γ/B- resonances), then it is clearly a family of non-overlapping resonances in A (i.e. γ-resonances). (In fact, it is a family of non-overlapping γ'-resonances for any $\gamma' : A \to \mathbf{C}$ such that $\gamma'/B = \gamma/B$.) But J may not belong to $ad(\gamma)$ even if it belongs to $ad(\gamma/B)$.

Lemma 7. Let $J \in ad(\gamma)$ and let $(c,d) \in J$. If (e,d) is a resonance, $e \neq c$, then $(e,d) \notin J$.

Proof. Suppose $(e,d) \in J$. Since (c,d) and (e,d) are not disjoint, one, (c,d) say, must be smaller than the other. Let (g,h) be the smallest resonance in J such that $g \prec c \prec h$. Since (g,h) is smaller than (e,d), we have $h \preceq d$. But (c,d) must be smaller than (g,h), so it follows that $d \preceq h$. Hence, $h = d$ and therefore $\gamma(J)(c) = \gamma(c) - \gamma(g) = 0$. This contradiction proves the lemma. ∎

Lemma 8. Assume $\gamma(b) \neq 0$. Let $c \in A = A(b)$ and suppose (c,b) is a resonance. Then

$$ad(\gamma/A\backslash]c,b]) \times ad(\gamma - \gamma(c)/]c,b[) \times \{\{(c,b)\}\}$$

is the set of all families in $ad(\gamma)$ which contain the resonance (c,b).

Proof. (c,b) is a resonance, $\{(c,b)\}$ is the family which consists of this single resonance, and $\{\{(c,b)\}\}$ is the set which consists of this single family.

The product of two sets of families is defined by identifying $(J,\tilde{J})$ with $J \cup \tilde{J}$. Let now $J = J_1 \cup J_2 \cup \{(c,b)\}$. If the right hand side is an element in the product, then clearly J is a family of non-overlapping γ-resonances containing (c,b). And if $J \in ad(\gamma)$ contains (c,b), then, by lemma 7, $J = J/A\backslash]c,b] \cup J/]c,b[\cup \{(c,b)\}$. Also the admissibility carries over from one side to the other since $\gamma(J)(x) = \gamma(J_1)(x)$ if $x \in A\backslash]c,b]$, $\gamma(J)(x) = (\gamma - \gamma(c))(J_2)(x)$ if $x \in]c,b[$, and $\gamma(J)(b) = \gamma(b)$. ∎

The following proposition describes inductive relations for the set of admissible families.

Proposition 1. Assume $\gamma(b) \neq 0$. Let $b_1,...,b_n$ be all the indices in $A \setminus \{b\}$ such that (b_j,b) is a resonance, and let

$$S_j = ad(\gamma/A\backslash]b_j,b]) \times ad(\gamma - \gamma(b_j)/]b_j,b[).$$

Then 1)

$$ad(\gamma) = \cup_{j=1}^{n}(S_j \times \{\{(b_j,b)\}\}) \cup ad(\gamma/A \setminus \{b\})$$

and

2) $S_j \subset ad(\gamma/A \setminus \{b\})$, and if $b_i \prec b_j$, then $S_i \cap S_j = \emptyset$.

Proof. In order to prove 1), we decompose $ad(\gamma)$ as $E_1 \cup ... \cup E_n \cup E_0$, where E_j is the set of all families that contain (b_j,b), and $E_0 = ad(\gamma/A \setminus \{b\})$. This decomposition is disjoint, by lemma 7, and the description of E_j follows from lemma 8. The first part of 2) is also obvious, and, in order to prove the second part, we let $J = J_1 \cup J_2 \in S_i$, $b_i \prec b_j$. Since $0 \neq (\gamma - \gamma(b_i))(J_2)(b_j)$ it follows that J_2 must contain a resonance (c,d) such that $c \prec b_j \prec d$ and $\gamma(b_j) \neq \gamma(c)$. But then (c,d) and (b_j,b) will overlap, so J can not belong to S_j. This proves 2). ∎

Of course, since A is a linear it always holds that either $b_i \prec b_j$ or $b_j \prec b_i$. Hence, the proposition has the obvious

Corollary. $\#ad(\gamma) \leq 2^{\#A-1}$.

An equivalence relation on $ad(\gamma)$.

From now on it is essential that $A = [a,b]$ is linear.

Definition A γ-resonance (c,d) is said to be *critical* if

$$|\gamma(x)| > 3 |\gamma(c)|$$

for all $x \in]c,d[$.

Notice that any two critical resonances are non-overlapping.

In a remark to Siegel's first lemma (lemma 4) we have explained why the critical resonances are important. From this point of view it would appear natural to replace the constant 3 by 1, but our argument requires a constant larger than 1.

Definition. Let J be an admissible family. We let $C'(\gamma, J)$ be the family of all γ-resonances in J together with all critical $(\gamma/A \setminus suppJ)$-resonances, and we let $C(\gamma, J)$ be the family of all maximal elements in $C'(\gamma, J)$.

$C(\gamma, J)$ is an admissible family of pairwise disjoint γ-resonances. Notice also that a critical γ-resonance which lies in $A \setminus suppJ$ is a critical $(\gamma/A \setminus suppJ)$-resonance, but there are critical $(\gamma/A \setminus suppJ)$-resonances which are not critical as γ-resonances.

Definition. We say that J and $\tilde{J}$ in $ad(\gamma)$ are *equivalent*, $J \sim_\gamma \tilde{J}$, if

$$C(\gamma, J) = C(\gamma, \tilde{J}).$$

Let $[J]_\gamma$ denote the equivalence class of J. $[J]_\gamma$ is *minimal* if $C(\gamma, J)$ is void, and it is *maximal* if $C(\gamma, J) = \{(a, b)\}$.

Notice that $C(\gamma, J)$ is void if and only if J is the trivial family, and there are no critical γ-resonances at all. Moreover, there is at most one minimal class in $ad(\gamma)$, and it consists precisely of the trivial family.

We shall now describe the structure of the maximal classes.

Proposition 2. Let $C(\gamma, J) = \{(c_i, d_i) : i \leq n\}$, and let $C_i = [c_i, d_i]$ and $\gamma_i = \gamma/C_i$. Then

$$[J]_\gamma \;=\; \prod_{1 \leq i \leq n} [J/C_i]_{\gamma_i}.$$

Proof. We first want to prove that $[J]_\gamma \subset \prod_{1 \leq i \leq n} [J/C_i]_{\gamma_i}$. This amounts to proving that if $\tilde{J} \sim_\gamma J$, then $\tilde{J}/C_i \sim_{\gamma_i} J/C_i$. Clearly $\tilde{J}/C_i \in ad(\gamma_i)$, so it suffices to show that

$$C(\gamma/C_i, \tilde{J}/C_i) \ni (c_i, d_i)$$

which is pretty obvious.

In order to prove that $[J]_\gamma \supset \prod_{1 \leq i \leq n} [J/C_i]_{\gamma_i}$, let $J_i \sim_{\gamma_i} J/C_i$. Then clearly $\tilde{J} = J_1 \cup ... \cup J_n \in ad(\gamma)$, and we must show that $\tilde{J} \sim_\gamma J$, i.e.

$$C(\gamma, \tilde{J}) = \{(c_i, d_i) : i \leq n\}$$

or, since there are no critical γ-resonances contained in $A \setminus \cup_{1 \leq i \leq n}]c_i, d_i[$,

$$C(\gamma, \tilde{J}) \supset \{(c_i, d_i) : i \leq n\}.$$

Hence, we must show that $(c_1, d_1) \in C(\gamma, \tilde{J})$. It is clear that either $(c_1, d_1) \in \tilde{J}$ or (c_1, d_1) is a critical $(\gamma/A \setminus supp\tilde{J})$-resonance. So the only way for (c_1, d_1) not to lie in $C(\gamma, \tilde{J})$ would be that there existed a critical $(\gamma/A \setminus supp\tilde{J})$-resonance (e, f) which contained (c_1, d_1). Since critical resonances are non-overlapping, it follows that (e, f) does not overlap any (c_i, d_i), and, hence, neither e nor f are contained in $]c_i, d_i[$.

Since (e, f) is not a critical $(\gamma/A \setminus suppJ)$-resonance, there must be an $x \in]e, f[\setminus suppJ$ such that

$$3 \mid \gamma(e) \mid \geq \mid \gamma(x) \mid .$$

This x must lie in $supp\tilde{J}$ since (e, f) is a critical $(\gamma/A \setminus supp\tilde{J})$-resonance, and, hence, $x \in]c_i, d_i[$ for some i, $i = 2$ say. Since (c_2, d_2) is critical for $(\gamma/A \setminus suppJ)$, it follows that

$$\mid \gamma(x) \mid > \; 3 \mid \gamma(c_2) \mid .$$

Finally, c_2 must lie in $]e,f[$, and since $c_2 \notin supp\tilde{J}$, we have

$$| \gamma(c_2) | > \; 3 \,| \gamma(e) | \,.$$

Hence,

$$3 \,| \gamma(e) | \geq | \gamma(x) | > \; 3 \,| \gamma(c_2) | > \; 9 \,| \gamma(e) |$$

which gives a contradiction. Hence, $(c_1, d_1) \in C(\gamma, \tilde{J})$. This proves the proposition. ∎

Remark. Let $\gamma' = \gamma/]a,b[$. If γ does not vanish at a or b, then $ad(\gamma') \subset ad(\gamma)$ and $[J]_{\gamma'} \subset [J]_\gamma$ for any $J \in ad(\gamma')$.

If $\gamma(a)$ and $\gamma(b)$ only are attained at a and b, respectively, these inclusions are equalities. But if this is not the case, then, not only $ad(\gamma') \neq ad(\gamma)$ but it may happen that $[J]_{\gamma'} \neq [J]_\gamma \cap ad(\gamma')$. For example if (a,b) and (c,d) are the only γ-resonances, and (a,b) is critical and (c,d) is not critical, then $[J_\emptyset]_{\gamma'} = \{J_\emptyset\}$ but $[J_\emptyset]_\gamma = ad(\gamma)$.

Hence, it is important to keep track on the domain of definition of γ, and it may be harmful to use γ to denote $\gamma/]a,b[$.

Since, by proposition 2, each class which is not minimal is a product of maximal classes, it will be important to understand the maximal classes. Suppose $\gamma(a) = \gamma(b)$ and let

$$S^1(\gamma) \;=\; \{J \in ad(\gamma - \gamma(a)/]a,b[) : | \gamma(x) | > 3 \,| \gamma(b) | \;\; for\ all\ x \in]a,b[\backslash suppJ\}$$

and

$$S^2(\gamma) \;=\; ad(\gamma - \gamma(a)/]a,b[) \setminus S^1(\gamma).$$

Lemma 9. $S^1(\gamma)$ is a saturated subset of both $ad(\gamma - \gamma(a)/]a,b[)$ and $ad(\gamma/]a,b[)$.

Proof. Clearly $S^1(\gamma)$ is a subset of $ad(\gamma')$, $\gamma' = \gamma/]a,b[$. In order to show that $S^1(\gamma)$ is saturated, we must prove that if $J_1 \sim_{\gamma'} J_2$ and $J_1 \in S^1(\gamma)$, then $J_2 \in S^1(\gamma)$. Suppose not. Then there exists $x \in]a,b[\backslash suppJ_2$ such that

$$3 \,| \gamma(a) | \geq | \gamma(x) | \,.$$

Since $J_1 \in S^1(\gamma)$ we have that $x \in suppJ_1$, and, hence, there is a resonance $(c,d) \in C(\gamma', J_1)$ such that $c \prec x \prec d$. Since $(c,d) \in C(\gamma', J_2)$, it follows that

$$| \gamma(x) | > \; 3 \,| \gamma(c) | \,.$$

Finally, since $c \notin suppJ_1$ we also have that

$$| \gamma(c) | > \; 3 \,| \gamma(a) | \,.$$

So we have proved that

$$3 \,| \gamma(a) | \geq | \gamma(x) | > \; 3 \,| \gamma(c) | > \; 9 \,| \gamma(a) |$$

which gives a contradiction. Hence, $J_2 \in S^1(\gamma)$ and $S^1(\gamma)$ is saturated in $ad(\gamma/]a,b[)$.

Let now $\gamma'' = (\gamma - \gamma(a))/]a,b[)$. In order to prove that $S^1(\gamma)$ is saturated in $ad(\gamma'')$, we must prove that if $J_1 \sim_{\gamma''} J_2$ and $J_1 \in S^1(\gamma)$, then $J_2 \in S^1(\gamma)$. We now argue as above. So suppose not. Then there exists $x \in]a,b[\backslash suppJ_2$ such that

$$3 \,| \gamma(a) | \geq | \gamma(x) | \,.$$

And there is a $c \notin suppJ_1$ such that

$$| \gamma(x) - \gamma(a) | > \; 3 \,| \gamma(c) - \gamma(a) |$$

and

$$| \gamma(c) | > \; 3 \,| \gamma(a) | \,.$$

So we have

$$\begin{aligned} 3 \mid \gamma(a) \mid \ &\geq \mid \gamma(x) \mid > \ 3 \mid \gamma(c) \mid -4 \mid \gamma(a) \mid \\ &\geq \ (3 - \frac{4}{3}) \mid \gamma(c) \mid = \ \frac{5}{3} \mid \gamma(c) \mid > \ 5 \mid \gamma(a) \mid \end{aligned}$$

which gives a contradiction. Hence, $J_2 \in S^1(\gamma)$ and $S^1(\gamma)$ is saturated in $ad(\gamma - \gamma(a)/]a, b[)$. ∎

Remark. It is only in the proof of the fact that $S^1(\gamma)$ is saturated in $ad(\gamma - \gamma(a)/]a, b[)$ that we need the factor 3 in the definition of a critical resonance. Indeed, one can replace 3 by any constant $d > 1 + \sqrt{2}$, but nothing smaller.

Proposition 3. Let $J \in ad(\gamma)$ and let $(c, d) \in C(\gamma, J)$. Then

$$[J/[c, d]]_\gamma = (S^1(\gamma/[c, d]) \times \{J_\emptyset, \{(c, d)\}\}) \cup (S^2(\gamma/[c, d]) \times \{\{(c, d)\}\}.$$

Proof. We can assume that $(c, d) = (a, b)$. Lemma 8 implies that $ad(\gamma - \gamma(a)/]a, b[) \times \{\{(a, b)\}\}$ is the set of all families $\tilde{J} \in ad(\gamma)$ which contains (a, b) and, hence, are equivalent to J. On the other hand, $S^1(\gamma)$ is precisely the set of all families $\tilde{J}$ which are equivalent to J but which does not contain (a, b). This proves the proposition. ∎

Generalization of Siegel's first lemma.

If $J \in ad(\gamma)$, then we define numbers $\rho(\gamma, J)(x)$ in the following way. We have to distinguish two cases: either it is the case that $\gamma(x) = \gamma('x)$ - $'x$ denotes the immediate predecessor of x - and there is a d such that $(x, d) \in J$, or this is not the case. Now we let

$$\underline{\rho(\gamma, J)(x)} \ = \ \begin{cases} \frac{1}{\gamma(J)(x)} & \text{in the second case} \\ \frac{1}{\gamma(J)(x)} - \sum_{y \in]x,d[\setminus supp(J/]x,d[)} \frac{1}{\gamma(J)(y)} & \text{in the first case.} \end{cases}$$

Let

$$\underline{\Phi(\gamma, \mathcal{D})} \ = \ \sum_{J \in \mathcal{D}} \prod_{x \in A} \rho(\gamma, J)(x)(-1)^{\# J}, \qquad \underline{\Phi(\gamma)} = \Phi(\gamma, ad(\gamma))$$

for any subset $\mathcal{D} \subset ad(\gamma)$, and let

$$\underline{\Phi'(\gamma)} = \frac{d}{dz}\Phi(\gamma_z)_{z=0}, \qquad \gamma_z = \gamma - z.$$

Definition. A short γ-resonance (c, d) is said to be *simple* if neither $('c, c)$ nor (d, d') are short γ-resonances.

Proposition 4. Suppose that $A = [a, b]$ and δ is a linear index set. Let $v : A \to \mathbf{Z}^\nu$ and let $\gamma = \gamma_{\delta,v}$. Assume that all short γ-resonances are simple.

Then

$$\mid \Phi(\gamma, [J]_\gamma) \mid \leq \ \#[J]_\gamma (2^{2\tau+6}K)^{\# A} \prod_{\substack{x \in A \\ v(x) \neq 0}} \mid v(x) \mid^{2\tau}$$

and, in particular,

$$\mid \Phi(\gamma) \mid \leq \ \# ad(\gamma) K (2^{2\tau+6}K)^{\# A} \prod_{\substack{x \in A \\ v(x) \neq 0}} \mid v(x) \mid^{2\tau} .$$

Moreover,

$$\mid \Phi'(\gamma) \mid \leq \ \# ad(\gamma) K (2^{3\tau+7}K)^{\# A} \prod_{\substack{x \in A \\ v(x) \neq 0}} \mid v(x) \mid^{3\tau} .$$

Remark. At the end of section VI we shall describe a simplified problem - linearization of a vector field on a torus. In that problem all short resonances are automatically compensated, so one can simply assume that there are no short resonances at all, in which case the expression for $\rho(\gamma, J)$ is much simpler since we always are in the second case.

In the Hamiltonian problem, however, not all short resonances are compensated for. Those that are not are always simple, because $\hat{F}_1(0)^2 = 0$, and it may therefore seem like they give rise to at most a "squaring" of a small divisor. But a simple short resonance may be followed by a (non-short) resonance giving rise to a triple repetition of a small divisor, and such a triple may repeat itself many times. The idea behind the particular formulation of $\rho(\gamma, J)$ in the first case is to compensate for such triple repetitions.

The basic compensations.

From now on we assume that all short γ-resonances are simple and $\gamma(a) \neq 0$.

Case 0.
Suppose that there are no resonances. Then

$$\Phi(\gamma) \;=\; \prod_{a \leq x \leq b} \gamma(x)^{-1}$$

if $\gamma \neq 0$. (If γ vanishes, then $\Phi(\gamma) = 0$.) In this case we can estimate $\Phi(\gamma)$ by Siegel's argument. (Lemma 10 describes a slightly stronger result which we will need.)

Lemma 10. Assume that all short γ-resonances are simple. Assume that γ never vanishes and that there are no critical γ-resonances. Then

$$\mid \prod_A \gamma(x)^{-1} \mid \;\leq\; (2^{2\tau+4}K)^{\#A} \prod_{\substack{x \in [a,b] \\ v(x) \neq 0}} \mid v(x) \mid^{2\tau} .$$

Proof. We shall proceed by induction on $\#A$ as in the proof of lemma 4. Notice that we can assume $K = 1$.

Choose x so that $\mid \gamma(x) \mid$ maximizes $\mid \gamma \mid$ on A, and take the smallest such x if there are several. If $\mid \gamma(x) \mid \geq 1$, then we can take away x and apply induction to $A \setminus \{x\}$, except if there is a resonance (c, d) such that x is the only element in $]c, d[$. But then $\mid \gamma(c) \mid \geq \frac{1}{3}$ and we can apply induction to $A \setminus \{c\}$.

So we can assume that $\mid \gamma \mid \leq 1$, and it suffices to show the estimate for $\mid \prod_A \gamma(x)^{-2} \mid$ under the assumption that $v \neq 0$.

Now this result just follows as in lemma 4. Choose x so that $\mid \gamma(x) \mid$ maximizes $\mid \gamma \mid$ on A and x is minimal. Induction works on $A \setminus \{x\}$ except if there is a resonance (c, d) such that x is the only element in $]c, d[$. But in that case we have

$$\mid \gamma(c) \mid \geq \frac{1}{4} max(\mid \langle v(c), \omega \rangle \mid, \mid \langle v(c'), \omega \rangle \mid)$$

and induction applies on $A \setminus \{c\}$. ∎

Case 1.
Suppose that there is only one resonance (a, b), and that it is non critical. Then

$$\Phi(\gamma) \;=\; \gamma(a)^{-1} \prod_{a<x<b} \gamma(x)^{-1} \gamma(b)^{-1} \;-\; \gamma(a)^{-1} \prod_{a<x<b} (\gamma(x) - \gamma(b))^{-1} \gamma(b)^{-1}$$

if $\gamma \neq 0$. (If γ vanishes, then we have only the second term.) Then we can apply Siegel's argument to each of the two terms. The first one is estimated by lemma 10 and the second one follows from lemma 11.

Lemma 11. Assume that $\gamma(a)=\gamma(b)\neq 0$ and that all short γ-resonances are simple. Assume also that $\gamma'=\gamma-\gamma(b)/]a,b[$ never vanishes, that there are no critical γ'-resonances, and that there is a $c\in]a,b[$ such that $|\gamma(c)|\leq 3|\gamma(b)|$. Then

$$|\gamma(a)\prod_{a<x<b}(\gamma(x)-\gamma(b))\gamma(b)|^{-1}\ \leq\ 4K^2(2^{2\tau+4}K)^{\#A-3}\prod_{\substack{x\in]a,b]\\ v(x)\neq 0}}|v(x)|^{2\tau}|\gamma(a)|^{-1}.$$

Proof. Since the result is trivial for $\#A=3$, we shall proceed by induction on $\#A$. We can assume $K=1$.

Arguing as in lemma 10, it suffices to prove the estimate for $|\gamma(a)\prod_{a<x<b}(\gamma(x)-\gamma(b))^2\gamma(b)|^{-1}$ under the assumption that $v\neq 0$.

Choose x such that $|\gamma'(x)|$ maximizes $|\gamma'|$ and x is minimal. If x is the unique index in $]a,b[$ such that $|\gamma(x)|\leq 3|\gamma(b)|$, then, for any $y\in]a,b[$,

$$|\gamma(x)-\gamma(b)|\geq|\gamma(y)-\gamma(b)|\geq\frac{1}{2}|\gamma(x)-\gamma(b)|$$

and, hence,

$$|\gamma(y)-\gamma(b)|\geq\frac{1}{4}max(|\langle v(y'),\omega\rangle|,|\langle v(y),\omega\rangle|).$$

And we can apply induction on $A\setminus\{'b\}$, for example.

If there is a resonance (e,f) such that x is the unique element in $]e,f[$, then $|\gamma'(x)|\leq 3|\gamma'(e)|$. Hence,

$$|\gamma'(e)|\geq\frac{1}{4}max(|\langle v(e),\omega\rangle|,|\langle v(e'),\omega\rangle|).$$

So in this case we can proceed by induction on $A\setminus\{e\}$.

If nothing of this is the case, then we do induction on $A\setminus\{x\}$. ∎

Case 2.

Suppose that there is only one resonance (a,b), and that it is critical. Then $\gamma\neq 0$, and $\Phi(\gamma)$ is the same as in case 1. But now each of the two terms is too big and it is only by considering their difference that we get a good estimate. We argue in the following way.

Let c minimize $|\gamma|$ on $]a,b[$. Then $|\gamma(c)|>3|\gamma(b)|$, and the function $f(z)=\prod_{a<x<b}\gamma_z(x)^{-1}$ is holomorphic for $z\in\{w\in\mathbf{C}:|w|<|\gamma(c)|\}$. By Cauchy estimates we get

$$|\Phi(\gamma)|=\frac{1}{|\gamma(a)|}\left|\frac{f(0)-f(\gamma(b))}{\gamma(b)}\right|\leq$$

$$\sup_{|z|\leq\frac{1}{2}|\gamma(c)|}|f(z)||\gamma(a)|^{-1}(\frac{1}{2}|\gamma(c)|-|\gamma(b)|)^{-1}\ \leq\ 2^{\#A-2}8|f(0)|\frac{1}{|\gamma(a)||\gamma(c)-\gamma(b)|}$$

where we have used that $|f(z)|\leq 2^{\#A-2}|f(0)|$ when $|z|<\frac{1}{2}|\gamma(c)|$, and that $|\gamma(c)-\gamma(b)|\leq 8(\frac{1}{2}|\gamma(c)|-|\gamma(b)|)$.

Hence, it suffices to estimate

$$\gamma(a)^{-1}\prod_{a<x<b}\gamma(x)^{-1}(\gamma(c)-\gamma(b))^{-1}$$

and this can be done by Siegel's argument.

Lemma 12. Assume that $\gamma(a)=\gamma(b)\neq 0$ and that all short γ-resonances are simple. Assume also that $\gamma'=\gamma/]a,b[$ never vanishes, that there are no critical γ'-resonances, and that $|\gamma(c)|>3|\gamma(b)|$, where c is determined so that $|\gamma(c)|=min\{|\gamma(y)|:a<y<b\}$. Then

$$|\gamma(a)\prod_{a<x<b}\gamma(x)(\gamma(c)-\gamma(b))|^{-1}\ \leq\ 4K^2(2^{2\tau+4}K)^{\#A-3}\prod_{\substack{x\in]a,b]\\ v(x)\neq 0}}|v(x)|^{2\tau}|\gamma(a)|^{-1}.$$

Proof. Since the result is trivial for $\#A = 3$, we shall proceed by induction on $\#A$. We can assume $K = 1$.

Arguing as in lemma 10 it suffices to prove the estimate for $\mid \gamma(a) \prod_{a<x<b} \gamma(x)^2 \gamma(b) \mid^{-1}$ under the assumption that $v \neq 0$.

Choose x so that $\mid \gamma(x) \mid$ maximizes $\mid \gamma \mid$ and x is minimal. If $x = c$, then, for all $y \in]a, b[$,

$$\mid \gamma(x) \mid = \mid \gamma(y) \mid$$

and, hence,

$$\mid \gamma(y) \mid \geq \frac{1}{2} max(\mid \langle v(y), \omega \rangle \mid, \mid \langle v(y'), \omega \rangle \mid).$$

And we can apply induction on $A \setminus \{'b\}$.

If there is a resonance (e, f) such that x is the unique element in $]e, f[$, then $\mid \gamma(x) \mid \leq 3 \mid \gamma(e) \mid$. Hence,

$$\mid \gamma(e) \mid \geq \frac{1}{4} max(\mid \langle v(e), \omega \rangle \mid, \mid \langle v(e'), \omega \rangle \mid).$$

So induction applies on $A \setminus \{e\}$.

If nothing of this is the case, then we do induction on $A \setminus \{x\}$. ▮

Case 3.

Suppose that there are only two resonances (a, b) and (a', b), and that (a', b) is non critical. Then

$$\Phi(\gamma) \;=\; \gamma(a)^{-1}\gamma(a')^{-1} \prod_{a'<x<b} \gamma(x)^{-1}\gamma(b)^{-1} \;-\; \gamma(a)^{-1}\gamma(a')^{-1} \prod_{a'<x<b} (\gamma(x) - \gamma(b))^{-1}\gamma(b)^{-1}$$
$$+\; \gamma(a)^{-1}(\sum_{a'<x<b} (\gamma(x) - \gamma(b))^{-1}) \prod_{a'<x<b} (\gamma(x) - \gamma(b))^{-1}\gamma(b)^{-1}$$

if $\gamma \neq 0$. (If γ vanishes, then we have only the second and the third term.)

Then we get estimates for each term separately by Siegel's argument. The first term is estimated by lemma 10 and the two others by lemma 13.

Lemma 13. Assume that $\gamma(a) = \gamma(a') = \gamma(b) \neq 0$ and that all short γ-resonances are simple. Assume also that $\gamma' = \gamma - \gamma(b)/]a', b[$ never vanishes, that there are no critical γ'-resonances, and that there is an index $c \in]a', b[$ such that $\mid \gamma(c) \mid \leq 3 \mid \gamma(b) \mid$. Then, for any $y \in]a', b[$,

$$\mid \gamma(a)(\gamma(y) - \gamma(b)) \prod_{a'<x<b} (\gamma(x) - \gamma(b))\gamma(b) \mid^{-1} \;\leq\; 4K^3(2^{2\tau+4}K)^{\#A-4} \prod_{\substack{x \in]a,b] \\ v(x) \neq 0}} \mid v(x) \mid^{2\tau} \mid \gamma(a) \mid^{-1} .$$

Proof. By exactly the same argument as in the proof of lemma 11 we see that it suffices to treat $\#A = 4$, in which case the estimates are trivial. ▮

Remark. Under the assumptions of lemma 13 we have $\mid \gamma(a') \mid^{-1} \leq 4 \mid \gamma(c) - \gamma(b) \mid^{-1}$. So we can replace $\gamma(y) - \gamma(b)$ by $4\gamma(a')$ in the above estimate.

Case 4.

Suppose that there are only two resonances (a, b) and (a', b), and that (a', b) is critical. Then $\gamma \neq 0$, and $\Phi(\gamma)$ is as in case 3. But now we have to proceed differently.

Let c minimize $\mid \gamma \mid$ on $]a', b[$. Then $\mid \gamma(c) \mid > 3 \mid \gamma(b) \mid$, and the function $f(z) = \prod_{a'<x<b} \gamma_z(x)^{-1}$ is holomorphic for $z \in \{w \in \mathbf{C} : \mid w \mid < \mid \gamma(c) \mid\}$. By Cauchy estimates we get

$$\mid \Phi(\gamma) \mid = \frac{1}{\mid \gamma(a)\gamma(a') \mid} \mid \frac{f(0) - f(\gamma(b))}{\gamma(b)} + f'(\gamma(b)) \mid \leq\; 2^{\#A-3}8^2 \mid f(0) \mid\mid \gamma(a) \mid^{-1} (\mid \gamma(c) - \gamma(b) \mid)^{-2}$$

where we have used that $\mid f(z) \mid \leq 2^{\#A-3} \mid f(0) \mid$ when $\mid z \mid < \frac{1}{2} \mid \gamma(c) \mid$, and that $\mid \gamma(c) - \gamma(b) \mid \leq 8(\frac{1}{2} \mid \gamma(c) \mid - \mid \gamma(b) \mid)$.

Hence, it suffices to estimate

$$\gamma(a)^{-1}\prod_{a'<x<b}\gamma(x)^{-1}(\gamma(c)-\gamma(b))^{-2}$$

and this can be done by Siegel's argument.

Lemma 14. Assume that $\gamma(a)=\gamma(a')=\gamma(b)\neq 0$ and that all short γ-resonances are simple. Assume also that $\gamma'=\gamma/]a',b[$ never vanishes, that there are no critical γ'-resonances, and that $\mid\gamma(c)\mid>3\mid\gamma(b)\mid$, where c is determined so that $\mid\gamma(c)\mid=min\{\mid\gamma(y)\mid: a'<y<b\}$. Then

$$\mid\gamma(a)\prod_{a'<x<b}\gamma(x)(\gamma(c)-\gamma(b))^2\mid^{-1}\;\leq\;4K^3(2^{2\tau+4}K)^{\#A-4}\prod_{\substack{x\in]a,b]\\ v(x)\neq 0}}\mid v(x)\mid^{2\tau}\mid\gamma(a)\mid^{-1}.$$

Proof. By exactly the same argument as in the proof of lemma 12 we see that it suffices to treat $\#A=4$, in which case the estimate is trivial. ∎

Remark. If we replace γ by $\gamma_z=\gamma-z$, then lemmas 10-14 remain valid for all complex numbers z such that $\mid z\mid<\frac{1}{2}min\{\mid\gamma(x)\mid: x\in A\}$ if we just multiply the constant by $2^{\#A}$. This is obvious since then $\mid\gamma(x)-z\mid\geq\frac{1}{2}\mid\gamma(x)\mid$.

Remark. The estimates are not optimal. In lemma 10, for example, the total weight of the exponents in $\prod\mid v(x)\mid^{2\tau}$ is $(\#A)2\tau$. But one can show, by the same type of argument, that there is an estimate with total weight $(\#A)\tau$. Moreover, the constant $2^{2\tau+4}$ can be replaced by $2^{\tau+2}$.

Proof of proposition 4.

We observe that the estimate is obvious for $\#A=1$, so we proceed by induction on $\#A$.

Suppose that $C(\gamma,J)$ is void. In this case $[J]_\gamma$ consists of only the trivial family, and there are no critical resonances. The estimate then follows from lemma 10.

Let's now consider the case when $C(\gamma,J)$ contains resonances. Let (c,d) be such that either $(c,d)\in C(\gamma,J)$ and $\gamma(c)\neq\gamma('c)$, or $(c',d)\in C(\gamma,J)$ and $\gamma(c)=\gamma(c')$ — we can assume that no (d,e) belongs to $C(\gamma,J)$. From proposition 2 it follows that

$$[J]_\gamma\;=\;[J/[c,d]]_{\gamma'}\times[J/A\backslash]c,d]]_{\gamma''}$$

where $\gamma'=\gamma/[c,d]$ and $\gamma''=\gamma/A\backslash]c,d]$. And then

$$\Phi(\gamma,[J]_\gamma)\;=\;\Phi(\gamma',[J/[c,d]]_{\gamma'})\times\gamma(c)\times\Phi(\gamma'',[J/A\backslash]c,d]]_{\gamma''}).$$

Hence, it suffices to prove that

$$\mid\Phi(\gamma,E)\mid\;\leq\;\#E(2^{2\tau+6}K)^{\#A-1}\prod_{\substack{x\in A-\{a\}\\ v(x)\neq 0}}\mid v(x)\mid^{2\tau}\mid\gamma(a)\mid^{-1}$$

in the following two cases: $E=[J]_\gamma$ and $(a,b)\in C(\gamma,J)$; $E=[J]_\gamma$ and $(a',b)\in C(\gamma,J)$.

By proposition 2 and 3 it suffices to prove this estimate in the following four cases:

$$E=[J]_{\gamma'-\gamma(b)}\times\{\{(a,b)\}\},\quad J\in S^2(\gamma)$$

$$E=[J]_{\gamma'}\times\{J_\emptyset,\{(a,b)\}\},\quad J\in S^1(\gamma)$$

where $\gamma'=\gamma/]a,b[$;

$$E=[J]_{\gamma'-\gamma(b)}\times\{\{(a',b)\}\},\quad J\in S^2(\gamma/[a',b])$$

$$E=[J]_{\gamma'}\times\{J_\emptyset,\{(a',b)\}\},\quad J\in S^1(\gamma/[a',b])$$

where $\gamma' = \gamma/]a', b[$ and $\gamma(a) = \gamma(a')$.

Case 1.

Since

$$\Phi(\gamma, [J]_{\gamma'-\gamma(b)} \times \{\{(a,b)\}\}) \;=\; -\,\gamma(a)^{-1}\Phi(\gamma' - \gamma(b), [J]_{\gamma'-\gamma(b)})\gamma(b)^{-1}$$

we can proceed by induction. It is therefore sufficient to consider the case when $C(\gamma' - \gamma(b), J)$ is void, and then the result follows from lemma 11 since $J \in S^2(\gamma)$.

Case 2.

We have

$$\Phi(\gamma, [J]_{\gamma'} \times \{J_\emptyset, \{(a,b)\}\}) \;=\; \gamma(a)^{-1}(\Phi(\gamma', [J]_{\gamma'}) - \Phi(\gamma' - \gamma(b), [J]_{\gamma'}))\gamma(b)^{-1}.$$

Now $z \to \Phi(\gamma' - z, [J]_{\gamma'})$ is holomorphic for

$$\mid z \mid < \mid \gamma(c) \mid = min\{\mid \gamma(x) \mid : x \in]a, b[\setminus suppJ\}$$

- this domain only depends on $[J]_{\gamma'}$. And since $3 \mid \gamma(a) \mid < \mid \gamma(c) \mid$, because $J \in S^1(\gamma)$, we get by Cauchy estimates

$$\mid \Phi(\gamma, [J]_{\gamma'} \times \{J_\emptyset, \{(a,b)\}\}) \mid \;\leq\; 2^{\#A-2}8 \mid \Phi(\gamma', [J]_{\gamma'}) \mid\mid \gamma(a)(\gamma(c) - \gamma(a)) \mid^{-1} .$$

We can now proceed by induction, and it suffices to consider the case when $C(\gamma', J)$ is void. Then the estimate follows from lemma 12 since $J \in S^1(\gamma)$.

Case 3.

We have

$$\Phi(\gamma, [J]_{\gamma'-\gamma(b)} \times \{\{(a', b)\}\}) \;=$$
$$-\,\gamma(a)^{-1}\gamma(a')^{-1}\Phi(\gamma' - \gamma(b), [J]_{\gamma'-\gamma(b)})\gamma(b)^{-1} \;+\; \gamma(a)^{-1}\Phi'(\gamma' - \gamma(b), [J]_{\gamma'-\gamma(b)})\gamma(b)^{-1}$$

where $\Phi'(\gamma' - \gamma(b), [J]_{\gamma'-\gamma(b)}) = \frac{d}{dz}\Phi(\gamma' - z, [J]_{\gamma'-\gamma(b)})/_{z=\gamma(b)}$.

For the first term we can proceed by induction, and it suffices to consider the case when $C(\gamma' - \gamma(b), J)$ is void. Then we can apply the remark to lemma 13.

The second term is a derivate so we can, using Cauchy estimates, estimate it by

$$2 \sup_{|z|<\frac{1}{2}|\gamma(c)-\gamma(b)|} \mid \gamma(a)^{-1}\Phi(\gamma'_z - \gamma(b), [J]_{\gamma'-\gamma(b)})(\gamma(c) - \gamma(b))^{-1}\gamma(b)^{-1} \mid \leq$$

$$2^{\#A-2} \mid \gamma(a)^{-1}\Phi(\gamma' - \gamma(b), [J]_{\gamma'-\gamma(b)})(\gamma(c) - \gamma(b))^{-1}\gamma(b)^{-1} \mid$$

where c is determined in order to minimize $\mid \gamma(x) - \gamma(b) \mid$ over $x \in]a', b[-suppJ$. (Notice that c only depends on $[J]_{\gamma'-\gamma(b)}$.)

Now we can proceed by induction, and therefore we only need to consider the particular case when $C(\gamma' - \gamma(b), J)$ is void. Then the estimate follows from lemma 13.

Case 4.

We have

$$\Phi(\gamma, [J]_{\gamma'} \times \{J_\emptyset, \{(a', b)\}\}) \;=$$
$$\gamma(a)^{-1}\gamma(a')^{-1}\gamma(b)^{-1}[\Phi(\gamma', [J]_{\gamma'}) - \Phi(\gamma' - \gamma(b), [J]_{\gamma'}) + \gamma(b)\Phi'(\gamma' - \gamma(b), [J]_{\gamma'})].$$

The bracket expression is the beginning of a power series expansion of a holomorphic function, so we can, by using Cauchy estimates, estimate it by

$$2^{\#A-3}8^2 \mid \Phi(\gamma', [J]_{\gamma'}) \mid\mid \gamma(b) \mid^2\mid \gamma(c) - \gamma(b) \mid^{-2}$$

where c is determined in order to minimize $\mid \gamma(x) \mid$ over $x \in]a', b[\setminus suppJ$. (Notice that c only depends on the class $[J]_{\gamma'}$.)

Now it suffices to consider the case when $C(\gamma', J)$ is void, and then we can apply lemma 14.

This concludes the proof of the first estimate proposition 4. In order to prove the second estimate we observe that the first estimate is valid for $\Phi(\gamma_z, [J]_\gamma)$, as long as $\mid z \mid < \frac{1}{2} \mid \gamma(x) \mid$ for all $x \notin suppJ$, if we just multiply the constant by $2^{\#A}$. And that $\Phi(\gamma_z, [J]_\gamma)$ is a holomorphic function in this domain. The second estimate then follows from the first by a Cauchy estimate. ∎

V GENERALIZATION OF SIEGEL'S THIRD LEMMA.

Resonances on index set.

Let $A = A(b)$ be a general index set, and let $\gamma : A \to \mathbf{C}$ be given. Then we can define γ-resonances and short γ-resonances exactly as in section IV, and all results from **Resonances on linear index sets** remain true. The proofs are exactly the same except for the estimate of $\#ad(\gamma)$, which in the linear case appears as an immediate corollary of proposition 1. This result we formulate here as

Lemma 15. $\#ad(\gamma) \ \leq \ 2^{\#A-1}$.

Proof. We refer to the decomposition of proposition 1. Suppose that b_1 is a smallest element among $b_1, ..., b_n$, and let a be the immediate predecessor of b in $[b_1, b]$. Then we can assume that $\gamma(a) \neq \gamma(b)$, since otherwise $ad(\gamma - \gamma(b)/]b_j, b[)$ would be void. We can now change A and γ to A' and $\gamma' : A' \to \mathbf{C}$ in the following way. A' is obtained from A by letting b_1 be the maximal element for the ordering, all other relations remain the same, and γ' is the same as γ except in the maximal element b_1, where we let $\gamma'(b_1) = \gamma'(a)$. A little reflection shows that $\#ad(\gamma') \geq \#ad(\gamma)$.

In this way we can proceed inductively. Hence we can assume that all b_j:s lie on a totally ordered subset, i.e. $b_i \prec b_j$ or $b_j \prec b_i$ for all i, j, and then lemma 15 follows immediately from proposition 1. ∎

In the case when $A = A_1 \cup ... \cup A_n$ is non-simple we clearly have $ad(\gamma) = ad(\gamma/A_1) \times ... \times ad(\gamma/A_n)$.

For $J \in ad(\gamma)$ we recall the definition of the numbers $\underline{\rho(\gamma, J)(x)}$. We have two cases: either it is the case that $\gamma(x) = \gamma('x)$ - a condition which requires that x has a unique immediate predecessor - and there exists a d such that $(x, d) \in J$, or this is not the case. (Notice that when $\gamma = \gamma_{\delta,v}$, then x has a unique immediate predecessor $'x$ with $\gamma(x) = \gamma('x)$ if and only if $(\delta(x), v(x)) = (1, 0)$.) Now $\rho(\gamma, J)(x)$ is defined by the formula of section IV. Also $\Phi(\gamma)$ and $\Phi'(\gamma)$ are defined as in section IV.

Generalization of Siegel's third lemma.

Proposition 5. Let $\delta \in \Delta(k)$ and $v \in \Gamma(k)$ and define $\gamma : A \to \mathbf{C}$ by $\gamma = \gamma_{\delta,v}$. Assume that all short γ-resonances are simple. Then

$$\mid \Phi(\gamma) \mid \leq \ \#ad(\gamma)(2^{14\tau+6}K)^{\#A} \prod_{\substack{x \in A \\ v(x) \neq 0}} \mid v(x) \mid^{8\tau} (2 \sum_{x \in A} \mid v(x) \mid)^{-6\tau}.$$

Proof. For any subset B of A, we let $\mid B \mid$ and $\overline{B}$ be defined as in lemma 6. Moreover, it will be convenient to use the same product convention:

$\prod_I c_i$ is the product of all c_i, $i \in I$, which are $\neq 0$, and it is 1 if no such c_i exist. In particular, it is 1 if $I = \emptyset$. In agreement with this, we let $\prod c$ be c if $c \neq 0$, and be 1 if $c = 0$.

Notice first that if $A = [a, b]$ is linear, then the estimate follows from proposition 4. This takes, in particular, care of the case $\#A = 1$, so we can proceed by induction on $\#A$.

Decomposition of $\Phi(\gamma)$.

Let T be the largest totally ordered subset $[a,b]$ of $A = A(b)$ such that

$$\mid A(x) \mid > \frac{1}{2} \mid A(x') \mid, \quad x \in [a,b[.$$

Let's say that a resonance (c,d) *cuts* T if $c \notin T$ and $d \in T$. We now define

$$E \;=\; \{J \in ad(\gamma) : no\ (c,d) \in J\ cuts\ T\}$$

and

$$E(c,d) \;=\; \{J \in ad(\gamma) : (c,d) \in J\ and\ no\ other\ resonance\ in\ J\ cuts\ [d,b]\}$$

for any γ-resonance (c,d) that cuts T.

Now we have

$$\Phi(\gamma) \;=\; \Phi(\gamma,E) + \sum \Phi(\gamma,E(c,d))$$

where summation runs over all resonances that cuts T, and

$$E = ad(\gamma/[a,b]) \times ad(\gamma/A \setminus [a,b])$$

$$E(c,d) = ad(\gamma/[d,b]) \times ad(\gamma - \gamma(c)/]c,d[) \times \{(c,d)\} \times ad(\gamma/A\setminus]c,b]).$$

We observe first that

$$\Phi(\gamma,E) = \Phi(\gamma/[a,b])\Phi(\gamma/A \setminus [a,b])$$

since $(\delta(a), v(a)) \neq (1,0)$. Secondly, if it is not the case that c has a unique predecessor $'c$ with $\gamma(c) = \gamma('c)$, then

$$\Phi(\gamma,E(c,d)) = \Phi(\gamma/[d,b])\Phi(\gamma - \gamma(c)/]c,d[)\Phi(\gamma/A\setminus]c,b]).$$

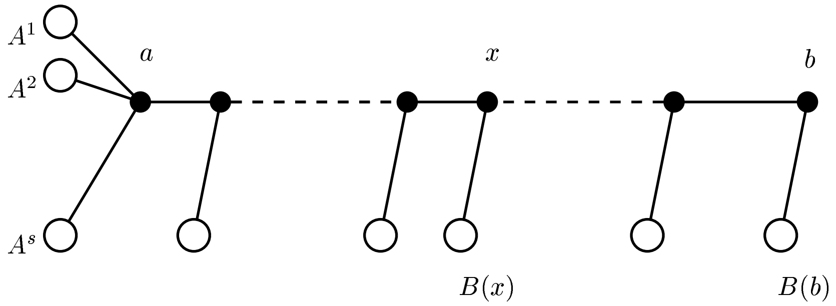
On the other hand, if this is the case, then

$$\Phi(\gamma,E(c,d)) = \Phi(\gamma/[d,b])\{\Phi(\gamma - \gamma(c)/]c,d[)\gamma(c)^{-1} - \Phi'(\gamma - \gamma(c)/]c,d[)\}\Phi(\gamma/A\setminus]'c,b]).$$

Hence, it suffices to treat these three cases. Since the appearance of the factor $\#ad$ is completely trivial, we shall simply suppress it in the estimations below.

First case: $\Phi(\gamma/[a,b])\Phi(\gamma/A \setminus [a,b])$.

We decompose A according to $[a,b]$:



Here $A(x) = \{x\} \cup A('x) \cup B(x)$ for each $x \in]a,b]$, where $'x$ is the predecessor of x in $[a,b]$, and $A(a) = \{a\} \cup A^s \cup ... \cup A^1$, $s \geq 0$, is the natural decomposition. Each A^i is a simple non-void index set, while the $B(x)$:s may be void, simple or non-simple.

We get from proposition 4,

$$\mid \Phi(\gamma/[a,b]) \mid \leq \; (2^{2\tau+6}K)^{\#[a,b]} \times \prod_{a \prec x \preceq b} \mid v(x) + \overline{B(x)} \mid^{2\tau} \times \mid \overline{A(a)} \mid^{2\tau}$$

and, by induction, we have

$$|\ \Phi(\gamma/B)\ |\ \leq\ (2^{14\tau+6}K)^{\#B}\times\prod_{x\in B}|\ v(x)\ |^{8\tau}\ \times(2\ |\ B\ |)^{-6\tau}$$

for any non-void $B = B(x)$, $x\ \in]a,b]$, or $B = A^i$.

We shall now take the product of all these estimates. We observe that $|\ v(x) + \overline{B(x)}\ |$ equals $|\ v(x)\ |$ when $B(x) = \emptyset$, and is less than $2\ |\ v(x)\ ||\ \overline{B(x)}\ |$ when $B(x)$ is non-void. But in the latter case we have a factor $2\ |\ B(x)\ |$ in the denominator. Hence we get

$$(2^{14\tau+6}K)^{\#A}\prod_{x\in A}|\ v(x)\ |^{8\tau}\ \times(2\ |\ A\ |)^{-6\tau}\times I$$

$$I\ =\ (2^{-12\tau})^{\#[a,b]}(\prod_{1\leq i\leq s}2\ |\ A^i\ |)^{-6\tau}(\prod|\ v(a)\ |)^{-8\tau}\ |\ \overline{A(a)}\ |^{2\tau}\ (2\ |\ A\ |)^{6\tau}$$

and we only need to prove that $I \leq 1$.

We have $|\ A(a)\ | > (\frac{1}{2})^{n-1}\ |\ A\ |$, $n = \#[a,b]$, and $|\ A^i\ | \leq \frac{1}{2}\ |\ A(a)\ |$.

Suppose first $|\ v(a)\ | \geq \frac{1}{2}\ |\ A(a)\ |$. If $|\ v(a)\ | = |\ A(a)\ |$, then we estimate I by

$$(2^{-12\tau})^{\#[a,b]}\ |\ v(a)\ |^{-8\tau}|\ \overline{A(a)}\ |^{2\tau}\ (2\ |\ A\ |)^{6\tau}$$

and the result follows. But if $|\ v(a)\ | < |\ A(a)\ |$, then $s \geq 1$ and we can estimate I by

$$(2^{-12\tau})^{\#[a,b]}2^{-6\tau}\ |\ v(a)\ |^{-8\tau}|\ \overline{A(a)}\ |^{2\tau}\ (2\ |\ A\ |)^{6\tau}$$

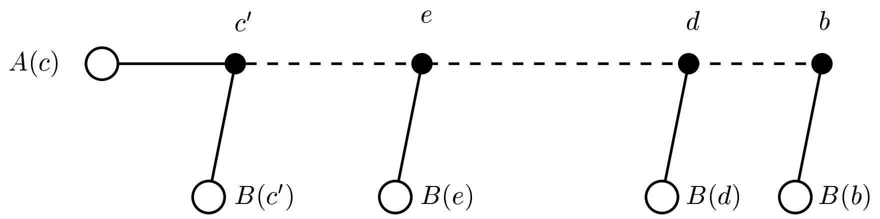
and the result follows again.

Suppose now instead that $|\ v(a)\ | < \frac{1}{2}\ |\ A(a)\ |$. Then $s \geq 2$, and the result follows by applying lemma 5 to

$$(\prod|\ v(a)\ |^2\prod_{1\leq i\leq s}(2\ |\ A^i\ |)^2)^{-3\tau}.$$

Second case: $\Phi(\gamma/[d,b])\Phi(\gamma-\gamma(c)/]c,d[)\Phi(\gamma/A\backslash]c,b])$.

Let e be the smallest element in $[a,b]$ such that $e\ \in]c,d]$, i.e. $[e,d] = [a,b]\cap]c,d]$. We shall decompose A according to $]c,b]$:



Here $A(x) = \{x\}\cup A('x)\cup B(x)$ for each $x\ \in]c,b]$, where $'x$ is the predecessor of x in $[c,b]$. The $B(x)$:s may be void or simple or non-simple.

We get by proposition 4,

$$|\ \Phi(\gamma/[d,b])\ |\ \leq\ (2^{2\tau+6}K)^{\#[d,b]}\times\prod_{d\prec x\preceq b}|\ v(x) + \overline{B(x)}\ |^{2\tau}\ \times|\ \overline{A(c)}\ |^{2\tau}$$

since $\overline{A(d)} = \overline{A(c)}$.

Also by proposition 4,

$$|\ \Phi(\gamma-\gamma(c)/]c,d[)\ |\ \leq\ (2^{2\tau+6}K)^{\#]c,d[}\times\prod_{c\prec x\preceq d}|\ v(x) + \overline{B(x)}\ |^{2\tau}\ .$$

If now $c \prec e \prec d$, then

$$v(e) + \overline{B(e)} \;=\; -\sum_{\substack{c \prec x \prec d \\ x \neq e}} v(x) + \overline{B(x)}$$

since $\overline{A(d)} = \overline{A(c)}$, and we have

$$\mid \Phi(\gamma - \gamma(c)/]c,d[) \mid \;\leq\; (2^{4\tau+6}K)^{\#]c,d[} \times \prod_{\substack{c \prec x \prec d \\ x \neq e}} \mid v(x) + \overline{B(x)} \mid^{4\tau} \times \mid v(d) + \overline{B(d)} \mid^{2\tau} .$$

So in both cases, $e \prec d$ and $e = d$, we get an estimate of $\Phi(\gamma/[d,b])\Phi(\gamma - \gamma(c)/]c,d[)$ which is independent of $v(e)$ and $\overline{B(e)}$.

By induction we have

$$\mid \Phi(\gamma/B) \mid \;\leq\; (2^{14\tau+6}K)^{\#B} \times \prod_{x \in B} \mid v(x) \mid^{8\tau} \times (2 \mid B \mid)^{-6\tau}$$

for any non-void $B = B(x)$, $x \in]c,b]$, or $B = A(c)$.

Putting these estimates together gives

$$(2^{14\tau+6}K)^{\#A} \prod_{x \in A} \mid v(x) \mid^{8\tau} \times (2 \mid A \mid)^{-6\tau} \times I$$

$$I = (2^{-10\tau})^{\#]c,b]} (\prod \mid v(e) \mid)^{-8\tau} (\prod 2 \mid B(e) \mid)^{-6\tau} (2 \mid A \mid)^{6\tau}$$

and we shall now show that $I \leq \frac{1}{2}$.

If $a \prec e$, then

$$\mid B(e) \mid \;\geq\; \frac{1}{2} \mid A(e) \mid \;\geq\; (\frac{1}{2})^n \mid A \mid, \quad n = \#[e,b] \leq \#]c,b]$$

and the the result follows.

If $e = a$, then $\mid A('e) \mid \leq \frac{1}{2} \mid A(e) \mid$, where $'e$ is the predecessor of e in $[c,b]$. But since $A(e) = \{e\} \cup A('e) \cup B(e)$, it follows that

$$\mid v(e) \mid + \mid B(e) \mid \;\geq\; \frac{1}{2} \mid A(e) \mid \;\geq\; (\frac{1}{2})^n \mid A \mid .$$

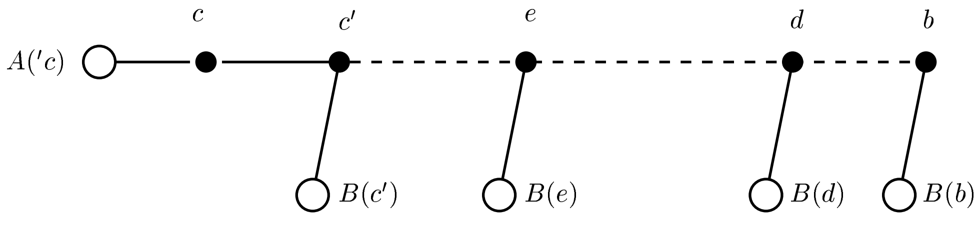
So if $B(e) = \emptyset$ or $v(e) = 0$, then

$$\mid v(e) \mid \;\geq\; (\frac{1}{2})^n \mid A \mid \quad \mathit{or} \quad \mid B(e) \mid \;\geq\; (\frac{1}{2})^n \mid A \mid$$

respectively, and in both cases the result follows. And otherwise we have $\mid v(e) \mid + \mid B(e) \mid \;\leq\; 2 \mid v(e) \mid \mid B(e) \mid$ and the result also follows.

Third case: $\Phi(\gamma/[d,b])\{\Phi(\gamma - \gamma(c)/]c,d[)\gamma(c)^{-1} - \Phi'(\gamma - \gamma(c)/]c,d[)\}\Phi(\gamma/A\backslash]'c,b])$.

We decompose A as in the second case:



The term is a sum of two pieces. The first one is estimated as in the second case just by replacing $A(c)$ by $A('c)$ and multiplying by $\mid \overline{A('c)} \mid^{\tau}$ because of the factor $\gamma(c)^{-1}$. For the second term we proceed in the same way. Hence, we get

$$\mid \Phi(\gamma/[d,b]) \mid \;\leq\; (2^{2\tau+6}K)^{\#[d,b]} \times \prod_{d \prec x \preceq b} \mid v(x) + \overline{B(x)} \mid^{2\tau} \times \mid \overline{A('c)} \mid^{2\tau}$$

since $\overline{A(d)} = \overline{A(c)} = \overline{A('c)}$, and

$$| \Phi'(\gamma - \gamma(c)/]c,d[) | \leq (2^{6\tau+7}K)^{\#]c,d[} \times \prod_{\substack{c \prec x \prec d \\ x \neq e}} | v(x) + \overline{B(x)} |^{6\tau} \times | v(d) + \overline{B(d)} |^{3\tau} .$$

By induction we have

$$| \Phi(\gamma/B) | \leq (2^{14\tau+6}K)^{\#B} \times \prod_{x \in B} | v(x) |^{8\tau} \times (2 \mid B \mid)^{-6\tau}$$

for any non-void $B = B(x)$, $x \in]a,b]$, or $B = A('c)$.

Putting these estimates together gives

$$(2^{14\tau+6}K)^{\#A} \prod_{x \in A} | v(x) |^{8\tau} \times (2 \mid A \mid)^{-6\tau} \times I$$

$$I = (2^{-8\tau+1})^{\#]c,b]} (\prod | v(e) |)^{-8\tau} (\prod 2 \mid B(e) \mid)^{-6\tau} (2 \mid A \mid)^{6\tau}.$$

We must now show that $I \leq \frac{1}{2}$, but this is just a repetition of the arguments in the second case.

This proves proposition 5. ∎

Remark. We have not tried to get the best constant and the estimates can certainly be ameliorated. We don't know which is the best exponent, but it seems likely that it is smaller than 8τ When there are no short resonances we have the same estimate with exponent 3τ instead of 8τ.

VI. ABSOLUTELY CONVERGENT SERIES.

We can now write down new series:

$$X(x) \sim \sum_{k \geq 1} \varepsilon^k (\sqrt{-1})^{-k} \sum_{\delta \in \Delta(k)} \sum_{v \in \Gamma(k)} \Lambda_1^*(\delta, v) \hat{F}(\delta, v) e^{\sqrt{-1}\langle v_1 + \ldots + v_k, x \rangle} \qquad (8)_1$$

$$C \sim \sum_{k \geq 1} \varepsilon^k (\sqrt{-1})^{-k} \sum_{\delta \in \Delta(k)} \sum_{v \in \Gamma(k)} \Lambda_2^*(\delta, v) \hat{F}(\delta, v) \qquad (8)_2$$

which we define in the following way:

$$\Lambda_1^*(\delta, v) = \begin{cases} \Phi(\gamma) = \sum_{J \in ad(\gamma)} \prod_{x \in A} \rho(\gamma, J)(x)(-1)^{\#J} & \text{if there are no short } \gamma\text{-res.} \\ & (c,d) \text{ with } \delta(d) \geq 2 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\Lambda_2^*(\delta, v) = \begin{cases} \Lambda_1^*(\delta, v/A \setminus \{b\}) & \text{if } \gamma(b) = 0 \\ \Lambda_2^*(\delta, v) = 0 & \text{otherwise} \end{cases}$$

where $\gamma = \gamma_{\delta,v}$ and $b = k$.

By proposition 5 and lemma 15 these series are absolutely convergent if, as in the Hamiltonian case, $\hat{F}_1(0)^2 = 0$. (This is so because $\hat{F}_1(0)^2 = 0$ implies that $\Lambda_1^*(\delta, v)\hat{F}(\delta, v) = 0$ unless all short $\gamma_{\delta,v}$-resonances are are simple.) And even though, in general, (8) does not converge to a solution of (*), we have not just replaced the absolutely divergent series (1) + (2) by *some* absolutely convergent series. The coefficient $\Lambda_1^*(\delta, v)$ is a sum of products of small divisors, one of which is $\Lambda_1(\delta, v)$. Hence, the series (8) contain (1)+(2)

- they are obtained from these latter by adding a large number of monomials $\hat{F}(\delta, v)$ with coefficients of type $\prod(\gamma(J)(x))^{-1}(-1)^{\# J}$.

So the series (8) will be a solution of (*) if and only if all these added terms sum up to 0. However, this is not always the case. In general we have

$$\partial X(x) \;=\; \varepsilon F(x, X(x)) - C - R(x)$$

with some $R(x) \neq 0$. But we are not so far away. We shall show that $R(x) = DX(x)$ for some matrix $D = D(\varepsilon)$, independent of x. Moreover, we shall show that, besides the mean value of X, there are more free parameters available, and by choosing these parameters appropriately one can, in the Hamiltonian case, make both C and D vanish.

Interpretation of the series.

Let

$$G(x, \varepsilon) \;=\; \sum_{k \geq 1} \varepsilon^k G_k(x) \;=\; \varepsilon F'_y(x, X(x)).$$

A simple calculation shows that $G_1 = F_1$ and

$$G_k(x) \;=\; \sum_{1 \leq j \leq k-1} \sum_{\substack{\iota_1 + \dots + \iota_j \\ = k-1}} (j+1) F_{j+1}(x)(X_{\iota_j}(x), ..., X_{\iota_1}(x))$$

for $k \geq 2$. Here $F_{j+1}(x)(X_{\iota_j}(x), ..., X_{\iota_1}(x))$ is a linear map, i.e. a matrix. Since F_{j+1} is symmetric, it doesn't matter in what order we evaluate it on the vectors $X_{\iota_j}(x), ..., X_{\iota_1}(x)$. This remark also explains the factor $(j+1)$.

For the Fourier coefficients we have

$$\hat{G}_k(w) \;=\; (\sqrt{-1})^{-k+1} \sum_{\delta \in \Delta(k)} \sum_{\substack{v \in \Gamma(k) \\ v_1 + \dots + v_k = w}} (\delta_k + 1) \hat{F}(\delta, v; k) \Lambda_1^*(\delta, v / \{1, ..., k-1\})$$

where $\underline{\hat{F}(\delta, v; j)}$ is the matrix defined by

$$\hat{F}(\delta, v; j) V = \hat{F}(\delta_k, v_k) ... \hat{F}(\delta_j + 1, v_j) V \hat{F}(\delta_{j-1}, v_{j-1}) ... \hat{F}(\delta_1, v_1)$$

for each vector V. For $\delta, v \in \Delta(k) \times \Gamma(k)$, the sets $A = \{1, ..., k\}$ and $A' = A \setminus \{k\}$ are index sets. A' has a natural decomposition $A^1 \cup ... \cup A^s$, and we understand by $\Lambda_1^*(\delta, v/A')$ the product $\prod \Lambda_1^*(\delta, v/A^i)$.

We now consider the equation

$$(**) \qquad \partial Z \;=\; (\varepsilon G - D) Z - \varepsilon Z \hat{G}_1(0), \quad < Z >= I$$

for a mapping $Z : \mathbf{T}^\nu \to gl(\mathbf{R}^\mu)$ and a constant matrix D. This equation has a formal solution

$$Z(x) \;\sim\; I + \sum_{k \geq 1} \varepsilon^k Z_k(x), \quad D \;\sim\; \sum_{k \geq 2} \varepsilon^k D_k$$

of the form

$$Z_k(x) \;=\; \sum_{1 \leq j \leq k} (\sqrt{-1})^{-j} \sum_{\substack{\iota_1 + \dots + \iota_j \\ = k}} \sum_{v \in \Gamma(j)} \Omega_1(\iota, v) \hat{G}_{\iota_j}(v_j) .... \hat{G}_{\iota_1}(v_1) e^{\sqrt{-1}\langle v_1 + ... v_j, x \rangle}$$

$$D_k \;=\; \sum_{1 \leq j \leq k} (\sqrt{-1})^{-j+1} \sum_{\substack{\iota_1 + \dots + \iota_j \\ = k}} \sum_{v \in \Gamma(j)} \Omega_2(\iota, v) \hat{G}_{\iota_j}(v_j) .... \hat{G}_{\iota_1}(v_1).$$

Let $\iota_1 + ... + \iota_j = k \geq j$ and $v \in \Gamma(j)$. We write $\{1, ..., j\} = [a, b]$, and we consider ι, v as a mapping on $[a, b]$. We let $\gamma(x) = \langle \sum_{y \leq x} v(y), \omega \rangle$.

Lemma 16. Suppose $\hat{G}_1(0)^2 = 0$. If $\hat{G}_{\iota_j}(v_j)...\hat{G}_{\iota_1}(v_1) \neq 0$, then

$$\Omega_1(\iota, v) = \begin{cases} \Phi(\gamma/[a,b]) & \text{if there are no short } \gamma\text{-res. } (c,d) \text{ with } \iota(d) \geq 2 \\ & \text{and if } (\iota(a), v(a)) \neq (1,0) \\ -\Phi'(\gamma/]a,b]) & \text{if there are no short } \gamma\text{-res. } (c,d) \text{ with } \iota(d) \geq 2 \\ & \text{and if } (\iota(a), v(a)) = (1,0) \\ 0 & \text{otherwise,} \end{cases}$$

$$\Omega_2(\iota, v) = \begin{cases} \Omega_1(\iota, v/[a,b[) & \text{if } \gamma(b) = 0 \text{ and } (\#[a,b] \geq 2 \text{ or } \iota(a) \geq 2) \\ 0 & \text{otherwise.} \end{cases}$$

(Here we shall understand $\Omega_1(\iota, v/\emptyset)$ to be 1.)

The reason why we must distinguish between those short γ-resonances (c, d) for which $\iota(d) \geq 2$ and those for which $\iota(d) \leq 1$ is that the former ones are automatically compensated for in the recurrence relations for the Ω_i:s by the presence of the term D in (**), while the latter ones are not compensated for because of the term $\varepsilon Z(x)\hat{G}_1(0)$. In $\Phi(\gamma)$, on the other hand, we have compensated for no short γ-resonances at all.

Proof. We have the following recurrence relations:

$$\begin{aligned} \gamma(b)\Omega_1(\iota, v) &= \Omega_1(\iota, v/[a,b[) \ - \ \Omega_2(\iota, v) \\ &\quad - \sum_{1 \leq j \leq n} \Omega_1(\iota, v/[a,b_j])\Omega_2(\iota, v/]b_j,b]) \ - \ \xi_a \Omega_1(\iota, v/]a,b]) \end{aligned}$$

where $b_1 < ... < b_n$, $b = b_{n+1}$, are all the indices in $[a, b[$ such that $\gamma(b_j) = \gamma(b)$, i.e. (b_j, b) are γ-resonances except eventually (b_n, b) which may be a short γ-resonance. ξ_a is 1 or 0 according to if $(\iota(a), v(a)) = (1, 0)$ or not.

From this we get that $\Omega_2(\iota, v) = \Omega_1(\iota, v/[a, b[)$ when $\gamma(b) = 0$, and 0 otherwise.

Now we turn to Ω_1. Clearly, Ω_1 is 0 when $\gamma(b) = 0$, so we assume that $\gamma(b) \neq 0$. Also, it follows by induction that $\Omega_1(\iota, v) = 0$ if there is a short γ-resonance (c, d) with $\iota(d) \geq 2$. Hence, we can assume that there are no such short resonances.

Consider now the decomposition $ad(\gamma) = \cup E_j$ given in proposition 1. We have

$$\Phi(\gamma) = \sum_{0 \leq j \leq n+1} \Phi(\gamma, E_j).$$

Suppose first that $\xi_a = 0$, i.e. $(\iota(a), v(a)) \neq (1, 0)$. We have the following relations:

$$\Phi(\gamma, E_{n+1}) = \frac{1}{\gamma(b)} \Phi(\gamma/[a,b[, ad(\gamma/[a,b[);$$

if $v(b_j) \neq 0$ and $v(b_{j+1}) \neq 0$, then

$$\Phi(\gamma, E_j) = -\Phi(\gamma/[a,b_j])\Phi(\gamma - \gamma(b)/]b_j,b[) \frac{1}{\gamma(b)};$$

if $v(b_j) = 0$, then E_{j-1} is void and $ad(\gamma/[a, b_j]) = ad(\gamma/[a, b_{j-1}])$, so

$$\Phi(\gamma, E_{j-1}) + \Phi(\gamma, E_j) =$$

$$-\Phi(\gamma, /[a,b_j])\Phi(\gamma - \gamma(b)/]b_j,b[) \frac{1}{\gamma(b)} + \Phi(\gamma, /[a,b_{j-1}])\Phi'(\gamma - \gamma(b)/]b_j,b[) \frac{1}{\gamma(b)}.$$

Now the statement follows by induction on $\#A$.

Suppose next that $\xi_a = 1$. Then we can assume $\xi_{a'} = 0$, because $\hat{G}_1(0)^2 = 0$, and using the preceding result we get

$$\begin{aligned}\gamma(b)\Omega_1(\iota, v) &= (\Omega_1(\iota, v/[a,b[) - \frac{1}{\gamma(b)}\Omega_1(\iota, v/]a,b[)) \\ &- \sum_{1\leq j\leq n}\{\Omega_1(\iota, v/[a,b_j]) - \frac{1}{\gamma(b)}\Omega_1(\iota, v/]a,b_j])\}\Omega_1(\iota, v/]b_j, b[)\end{aligned}$$

from which we immediately deduce the result. ∎

We can now expand Z and D in the monomials $\hat{F}(\delta, v; j)$. Then we get

$$Z_k = (\sqrt{-1})^{-k}\sum_{1\leq j\leq k}\sum_{\delta\in\Delta(k)}\sum_{v\in\Gamma(k)}\Lambda_3^*(\delta, v; j)(\delta_j + 1)\hat{F}(\delta, v; j)e^{\sqrt{-1}<v_1+\ldots+v_k, x>} \qquad (9)_1$$

$$D_k = (\sqrt{-1})^{-k+1}\sum_{1\leq j\leq k}\sum_{\delta\in\Delta(k)}\sum_{v\in\Gamma(k)}\Lambda_4^*(\delta, v; j)(\delta_j + 1)\hat{F}(\delta, v; j) \qquad (9)_2$$

where

$$\Lambda_3^*(\delta, v; a) = \begin{cases}\Phi(\gamma/[a,b])\Phi(\gamma/A\setminus[a,b]) & \text{if there are no short } \gamma\text{-res. } (c,d) \text{ with } \delta(d)\geq 2 \\ & \text{and if } (\delta(a), v(a)) \neq (0,0) \\ -\Phi'(\gamma/]a,b])\Phi(\gamma/A\setminus[a,b]) & \text{if there are no short } \gamma\text{-res. } (c,d) \text{ with } \delta(d)\geq 2 \\ & \text{and if } (\delta(a), v(a)) = (0,0) \\ 0 & \text{otherwise}\end{cases}$$

and

$$\Lambda_4^*(\delta, v; a) = \begin{cases}\Lambda_3^*(\delta, v/A\setminus\{b\}; a) & \text{if } \gamma(b) = 0 \text{ and } \#A\geq 2 \\ 0 & \text{otherwise.}\end{cases}$$

Consider now the the equations

$$(***) \qquad \begin{cases}\partial X(x) = \varepsilon F(x, X(x)) - C - DX(x), & < X >= 0 \\ \partial Z(x) = (\varepsilon F_y'(x, X(x)) - D)Z(x) - \varepsilon Z(x)\hat{F}_1(0), & < Z >= I.\end{cases}$$

Proposition 6. If $\hat{F}_1(0)^2 = 0$, then the series (8) + (9) satisfies (***).

Proof. The second equation is fulfilled by the very definition of (9), so we only need to consider the first one.

The recurrence relations are:

$$\gamma(b)\Lambda_1^*(\delta, v) = \Lambda_1^*(\delta, v/A\setminus\{b\}) - \Lambda_2^*(\delta, v) - \sum_{1\leq j\leq n}\Lambda_4^*(\delta, v/A\setminus A(b_j); b_j')\Lambda_1^*(\delta, v/A(b_j))$$

where $b_1, ..., b_n$ are all the indices in $A\setminus\{b\}$ such that $\gamma(b_j) = \gamma(b)$.

These relations are certainly fulfilled if $\gamma(b) = 0$, so we assume that $\gamma(b) \neq 0$. If there is a short γ-resonance (c,d) with $\delta(d)\geq 2$, then everything vanishes. So we can assume that there are no such short γ-resonances. Moreover, if, for example, (b_n, b) is a short γ-resonance, then the relations are also fulfilled. Hence, we can assume that all (b_j, b) are γ-resonances.

Under these assumptions we must show that

$$\gamma(b)\Phi(\gamma) = \Phi(\gamma/A\setminus\{b\}) - \sum_{1\leq j\leq n}\tilde{\Phi}(\gamma - \gamma(b)/]b_j, b[)\Phi(\gamma/A(b_j))\Phi(\gamma/A\setminus[b_j, b])$$

where

$$\tilde{\Phi}(\gamma - \gamma(b)/]b_j, b[) = \begin{cases}\Phi(\gamma - \gamma(b)/]b_j, b[) & \text{if } \gamma(b_j') \neq \gamma(b) \\ -\Phi'(\gamma - \gamma(b)/]b_j', b[) & \text{if } \gamma(b_j') = \gamma(b).\end{cases}$$

This is a trivial verification, of the same type as in the proof of lemma 16, which we leave to the reader.

∎

Hence, when $\hat{F}_1(0)^2 = 0$ it follows from proposition 5 and lemma 15 that the series (8)+(9) converge to a solution of (***). In particular, the series (8) converge to a solution of

$$\partial X(x) = \varepsilon F(x, X(x)) - C - DX(x), \quad < X >= 0$$

for some constant vector $C = C_\varepsilon$ and some constant matrix $D = D_\varepsilon$. C_ε and D_ε are analytic in ε and vanish at $\varepsilon = 0$, and the first derivative of D_ε also vanishes at $\varepsilon = 0$.

The solution X depends on the choice of mean value $< X >$ which we can consider as free parameters. But there are more parameters. In fact, the series (8) also solve

$$\partial X(x) = \varepsilon\{F(x, X(x)) + \varepsilon N X(x) - \varepsilon \hat{F}(1,0)X(x)\} - C - DX(x), \quad < X >= 0$$

for some constant vector $C = C_\varepsilon$ and some constant matrix $D = D_\varepsilon$, and for all constant matrices N such that $N^2 = 0$. C_ε and D_ε are analytic in N and ε and vanish at $\varepsilon = 0$, and now $D = \varepsilon(N - \hat{F}(1,0))$ modulo ε^2. This means that (8) is independent of the mean value $< \frac{\partial F}{\partial y}(.,0) >$, which we are free to vary as we like. Or, equivalently, we can consider $\hat{F}(1,0)$ are free parameters in the series (8), only subject to the condition $\hat{F}(1,0)^2 = 0$, which can be chosen independently of the mean value $< \frac{\partial F}{\partial y}(.,0) >$.

"Killing the constants".

In the Hamiltonian case, we are now confronted with the problem of killing all constants C and D. As parameters we have, as in the case discussed in section II, the mean values $< X >= (\lambda_1, \lambda_2)$, or rather just λ_2, since all the constants are independent of λ_1. But now we also have $\hat{F}(1,0)$. These parameters are sufficient to make both C and D vanish.

Consider
$$\begin{cases} \partial X(x) = \varepsilon J h'((x,\lambda_2) + X(x)) - C - DX(x), & < X >= 0 \\ \partial Z(x) = (\varepsilon J h''((x,\lambda_2) + X(x)) - D)Z(x) - \varepsilon Z(x) N, & < Z >= I. \end{cases}$$

where $N = \begin{pmatrix} 0 & M \\ 0 & 0 \end{pmatrix}$, M symmetric. By replacing D by $D + \varepsilon < Jh''(\ ,\lambda_2) > -\varepsilon N$ we can bring this equation on the form (***), so it has a unique analytic solution X, Z, C, D, which depends analytically on λ_2 and M. We shall now show that if $det < Q > \neq 0$, then λ_2 and M can be determined so that $C = 0$ and $D = 0$.

If we write
$$D = \begin{pmatrix} D^1 & D^2 \\ D^3 & D^4 \end{pmatrix}, \quad C = \begin{pmatrix} C^1 \\ C^2 \end{pmatrix}$$

and if we observe that
$$D = \varepsilon \begin{pmatrix} 0 & < Q > -M \\ 0 & 0 \end{pmatrix} \quad C = \varepsilon \begin{pmatrix} < Q > \lambda_2 \\ 0 \end{pmatrix}$$

modulo ε^2, then we can solve
$$D^2 + (D^2)^* = 0, \quad C^1 = 0 \tag{10}$$
uniquely for $\lambda_2 = \lambda_2(\varepsilon)$, $M = M(\varepsilon)$, since $det < Q > \neq 0$. A priori, this kills only a part of the constants, but in fact all C and D vanish for these values of λ_2 and M.

Lemma 17. $C = 0$ and $D = 0$ for $\lambda_2 = \lambda_2(\varepsilon)$, $M = M(\varepsilon)$.

Proof. In order to show this, we proceed as in lemma 1 and use the result of Poincaré which we have formulated there.

Differentiating the equality given there once and letting $y = 0$, we get for

$$\tilde{X}(x) = \Phi(x,0) - \begin{pmatrix} 0 \\ < \Phi^2(\ ,0) > \end{pmatrix}$$

that

$$\partial\tilde X(x)=\varepsilon Jh'((x,\lambda_2+<\Phi^2(\ ,0)>)+\tilde X(x))-\varepsilon Jf'(0)-\varepsilon\Phi'(x,0)Jf'(0),\qquad <\tilde X>=0$$

where $f'(y)$ is the gradient of f considered as a function of x and y, i.e. $f'(y)=\begin{pmatrix}0\\ \frac{\partial f}{\partial y}(y)\end{pmatrix}$. Differentiating once more and letting $y=0$ we get for $\tilde Z(x)=(I+\Phi'(x,0))\begin{pmatrix}I & 0\\ 0 & W\end{pmatrix}^{-1}$, $W=I+<\frac{\partial\Phi^2}{\partial y}(\ ,0)>$,

$$\partial\tilde Z(x)=\varepsilon Jh''((x,\lambda_2+<\Phi^2(\ ,0)>)+\tilde X(x))\tilde Z(x)-\varepsilon\tilde Z(x)Jf''(0)\begin{pmatrix}I & 0\\ 0 & W\end{pmatrix}^{-1}-\varepsilon\Phi''(x,0)Jf'(0),$$

$<\tilde Z>=I$.

Now $\frac{\partial f}{\partial y}(0)=<Q>\lambda+\mathbf{O}(\varepsilon)$, and, since $<Q>$ is non-singular, we can define a unique formal series $\tilde\lambda_2(\varepsilon)$ such that $f'(0)=0$ for $\lambda_2=\tilde\lambda_2(\varepsilon)$. And for this choice of λ_2 we let $M=\tilde M(\varepsilon)=\frac{\partial^2 f}{\partial y^2}(0)W^{-1}$.

With λ_2 and M so chosen, we get a formal solution $\tilde X,\ \tilde C=0,\ \tilde Z,\ \tilde D=0$ of the equation

$$\begin{cases}\partial\tilde X(x)=\varepsilon Jh'((x,\tilde\lambda_2+<\Phi^2(\ ,0)>)+\tilde X(x)), & <\tilde X>=0\\ \partial\tilde Z(x)=\varepsilon Jh''((x,\tilde\lambda_2+<\Phi^2(\ ,0)>)+\tilde X(x))\tilde Z(x)-\varepsilon\tilde Z(x)\tilde N, & <\tilde Z>=I.\end{cases}$$

This formal equation has a unique formal solution. But X,C,Z,D, for $\lambda_2=\tilde\lambda_2(\varepsilon)$ and $M=\tilde M(\varepsilon)$, is also a formal solution to this equation, so we conclude that

$$X=\tilde X,\ C=\tilde C,\ Z=\tilde Z,\ D=\tilde D.$$

Hence

$$C=0\quad and\quad D=0$$

for $\lambda_2=\tilde\lambda_2(\varepsilon)+<\Phi^2(\ ,0)>$ and $M=\tilde M(\varepsilon)$.

This implies that the formal series $\tilde\lambda_2(\varepsilon)+<\Phi^2(\ ,0)>$ and $\tilde M(\varepsilon)$ formally solve the equations (10). By the uniqueness of the solution of (10), it follows that these series converge and that

$$\tilde M(\varepsilon)=M(\varepsilon)\quad \tilde\lambda_2(\varepsilon)=\lambda_2(\varepsilon)-<\Phi^2(\ ,0)>.$$

This proves that C and D vanish for λ_2 and M chosen according to (10). ∎

We can now summarize our result in the following theorem, where we refer to the formulation of the Hamiltonian problem in the Introduction.

Theorem.

Let $F(x,y,\lambda_2)=Jh'((x,\lambda_2)+y)=\sum_{j\geq 0}F_j(x,\lambda_2)(y)^j$, and let $\hat F(j,w)$ be the w:th Fourier coefficient of F_j if $(j,w)\neq(1,0)$, and let $\hat F(1,0)=\begin{pmatrix}0 & M\\ 0 & 0\end{pmatrix}$, M symmetric, be arbitrary parameters.

Then the series

$$\sum_{k\geq 1}\varepsilon^k(\sqrt{-1})^{-k}\sum_{\delta\in\Delta(k)}\sum_{v\in\Gamma(k)}\Lambda_1^*(\delta,v)\hat F(\delta,v)e^{\sqrt{-1}\langle v_1+\ldots+v_k,x\rangle}$$

where

$$\Lambda_1^*(\delta,v)=\begin{cases}\sum_{J\in ad(\gamma)}\prod_{x\in A}\rho(\gamma,J)(x)(-1)^{\#J} & \text{if there are no short } \gamma\text{-res. } (c,d) \text{ with } \delta(d)\geq 2\\ 0 & \text{otherwise}\end{cases}$$

$$\rho(\gamma,J)(x)=\begin{cases}\frac{1}{\gamma(J)(x)}-\sum_{y\in]x,d[\setminus supp(J/]x,d[)}\frac{1}{\gamma(J)(y)} & \text{if } (\delta(x),v(x))=(1,0) \text{ and there is a } \gamma\text{-res. } (x,d)\in J\\ \frac{1}{\gamma(J)(x)} & \text{otherwise}\end{cases}$$

$(A = \{1, ..., k\},\ \gamma = \gamma_{\delta,v})$, converges to a solution of

$$\partial X(x) = \varepsilon J h'((x, \lambda_2) + X(x)) - C - DX(x), \quad < X >= 0$$

for some constant vector C and some constant matrix D and for all M and λ_2. (C and D are uniquely determined by equations (***).)

The convergence is uniform since

$$| \Lambda_1^*(\delta, v) | \leq \ (2^{14\tau+7}K)^{\# A} \prod_{v(x)\neq 0} | v(x) |^{8\tau} .$$

Both X, C and D are analytic functions of x, ε, λ_2 and M, and

$$C = \begin{pmatrix} C^1 \\ C^2 \end{pmatrix} = \varepsilon \begin{pmatrix} < Q > \lambda_2 \\ 0 \end{pmatrix}$$

$$D = \begin{pmatrix} D^1 & D^2 \\ D^3 & D^4 \end{pmatrix} = \varepsilon \begin{pmatrix} 0 & < Q > -M \\ 0 & 0 \end{pmatrix}$$

modulo ε^2.

Moreover, if $\frac{\partial^2 h}{\partial y^2}(x, 0) = Q(x)$ has a non-singular mean value, i.e. if $det < Q > \neq 0$, and if λ_2 and M are determined by

$$D^2 + (D^2)^* = 0, \quad C^1 = 0$$

then both C and D vanish, and X is a solution of

$$\partial X(x) = \varepsilon J h'((x, \lambda_2) + X(x)), \quad < X >= 0.$$

Other examples.

Example 1. The problem of linearizing a vector field on the torus $\mathbf{T}^\nu$ can also be put on this form.

Let $F(x, y) = G(x + y)$. Then the diffeomorphism $x \mapsto x + X(x)$ conjugates the perturbed vector field $\omega + \varepsilon G(x) - C$ with the linear vector field ω if and only if X and C satisfy equation (*). The formal solution (1) has an expansion (2) which is absolutely divergent. That, despite this, the formal solution is convergent was first shown by Arnold using KAM-technique [13].

Let's now consider the formal solution of (***). Since $\hat{F}_1(0) = \hat{G}'(0) = 0$ this solution is convergent, and we obtain for X'

$$\begin{aligned} \partial(I + X'(x)) &= \varepsilon G'(x + X(x))(I + X'(x)) - DX'(x) \\ &= (\varepsilon G'(x + X(x)) - D)(I + X'(x)) + D \end{aligned}$$

and for $W = Z^{-1}$

$$\partial W(x) = -W(x)(\varepsilon G'(x + X(x)) - D).$$

Hence,

$$\partial(W(x)(I + X'(x))) = -W(x)D.$$

This implies that $< W > D = 0$, and, since $W(x) = I + \mathbf{O}(\varepsilon)$, that $D = 0$.

Hence, the absolutely convergent series given by (8) solve this conjugation problem.

This case is much easier than the Hamiltonian case since $\hat{F}_1(0) = 0$ and, therefore, all short resonances are compensated for by the recurrence relations. As a consequence, one can assume that there are no short resonances at all, and this simplifies the estimations of the coefficients a lot. Also, in this case the estimates are much better:

$$| \Lambda_1^*(\delta, v) | \leq \ (2^{4\tau+4}K)^k \prod_{v(x)\neq 0} | v(x) |^{3\tau}$$

for all $\delta \in \Delta(k)$ and $v \in \Gamma(k)$.

Example 2. The simplest and most illuminating example of compensations of signs, however, is just the linear matrix equation

$$\partial Z(x) \;=\; (\varepsilon F_1(x) - D)Z(x), \qquad < Z > \;=\; I.$$

This equation has a formal solution which is convergent for ε sufficiently small, in contrast to the equation

$$\partial Z(x) \;=\; \varepsilon F_1(x)Z(x) - D, \qquad < Z > \;=\; I$$

which also has a formal solution, but which in general is divergent.

REFERENCES

[1] H. Poincaré, Les méthodes nouvelles de la mécanique céleste, vol. II, Paris (1892)

[2] A. N. Kolmogorov, The general theory of dynamical systems and classical mechanics, Proc. of the 1954 Intern. Congr. of Math., in R. Abraham, J. E. Marsden, Foundations of mechanics, Benjamin (1978)

[3] V. I. Arnold, Proof of A.N. Kolmogorov's theorem on the preservation of quasi-periodic motions under small perturbations of the Hamiltonian, Russ. Math. Surv. 18 (1962)

[4] J. Moser, Convergent series expansions for quasi periodic motions, Math. Ann. 169 (1967)

[5] C. L. Siegel, Iterations of analytic functions, Ann. Math. 43 (1942)

[6] L. H. Eliasson, Compensation of signs in a small divisor problem, in P. Dazord, N. Desolneux-Moulis, J.-M. Morvan, Aspects dynamiques et topologiques des groupes infinis de transformation de la mécanique, Hermann (1987)

[7] W. M. Schmidt, Diophantine Approximation, LNM 785, Springer (1980)

[8] C. L. Siegel, Vorlesungen über Himmelsmechanik, Springer (1956)

[9] S. Sternberg, Infinite Lie groups and the Formal Aspects of Dynamical Systems, J. Math. Mech. 10 (1961)

[10] C. L. Siegel, Über die Normalform analytischer Differentialgleichungen in der Nähe einer Gleichgewichtslösung, Nachr. Akad. Wiss. Göttingen, math.-phys. Kl. (1952)

[11] J. Pöschel, On invariant manifolds of complex analytic mappings near fixed points, Expo. Math. 4 (1986)

[12] A. D. Brjuno, Analytic form of differential equations, Trans. Moscow Math. Soc. 25 (1971)

[13] V. I. Arnold, Small denominators I. Mappings of the circumpherence onto itself, AMS Transl., series 2, 46 (1965)

Postscript 1996. The large class of compensations that "explains" what KAM-theory proves — namely that formal solutions for quasi-periodic motions converge — was found in the years 1985-88. The basic idea was presented in reference [6] and the whole analysis, applied to Hamiltonian systems was given in Report $2-88$ of the Department of Mathematics, University of Stockholm. This result was presented at several conferences these years, part of it was published in the book Analysis etc (P. Rabinowitz and E. Zehnder (eds.), Academic Press, 1990), and an application of these ideas to a conjecture by Gallavotti was published in the proceedings of a conference on Nonlinear Dynamics in Bologna (G. Turchetti (ed.), World Scientific, 1988). Since then there have been many improvements, simplifications and applications of these ideas, in particular by Gallavotti and collaborators, by Chiercha and Falcolini, and by Ecalle and Vallet.

The original report itself has, for different reasons, never been published. We do believe however that it — with its somewhat semi-direct approach — still may have some interest, besides an obvious historical one. The present article differs from the report only in minor respects: several misprints have been eliminated; some smaller errors, in particular in the proof of lemma 1, have been corrected: a few short explanations have been added.

We like to express our gratitude to Prof. G. Gallavotti for his interest and kind insistence in the publication of this work.