ISAS - INTERNATIONAL SCHOOL
FOR ADVANCED STUDIES

THESIS

SUBMITTED FOR THE DEGREE OF

MAGISTER PHILOSOPHIAE

LIE TRANSFORM TECHNIQUES AND NEKHOROSHEV-LIKE EXPONENTIAL ESTIMATES IN HAMILTONIAN PERTURBATION THEORY

Candidate: Francesco Fasso'

Supervisor: Prof. Giancarlo Benettin

Academic Year 1985-1986

TRIESTE

SISSA - SCUOLA
INTERNAZIONALE
SUPERIORE
STUDI AVANZATI
TRIESTE
Strada Costiera 11

CONTENTS

INTRODUCTION

Classical perturbation theory deals with nearly integrable Hamiltonian systems, described by a Hamiltonian which (typically) has the form

$$(1) \quad H(p,q,\varepsilon) = h(p) + \varepsilon f(p,q)$$

where p,q are action-angle variables for the unperturbed, integrable Hamiltonian described by h and ε is a "small" parameter. A basic information (since the q's are angles and the p's are constant in the unperturbed system) is to get an estimate of the "stability time" of the perturbed system, namely a lower bound $T(\varepsilon)$ of the times for which the actions remain "close" to their initial values, in the sense that it can be assured a bound $\|p(t)-p(0)\| \le \varepsilon^{a}$ for all $|t| \le T(\varepsilon)$, a being a positive constant.

Nekhoroshev theorem [38,39] (see also [9,11]) is a very strong result in this direction. Basically, it states that, for a large class of nearly integrable Hamiltonians, the stability time is longer than any power of $1/\varepsilon$; more precisely, for all the motions of the perturbed system (actually, with the exclusion of those motions which start "too near" the boundary of the domain on which the Hamiltonian is defined), and provided that ε is sufficiently small, one has a uniform estimate of the form

$$(2) \quad \|p(t)-p(0)\| \leq \mathcal{P}\,\varepsilon^{a}$$

for all

$$(3) \quad |t| \leq \mathcal{T}\, e^{(1/\varepsilon)^{b}}$$

where a, b, $\mathcal{P}$, $\mathcal{T}$ are positive numbers and $\|\cdot\|$ is a suitable norm. Nekhoroshev has proven this result under the hypothesis that $H=h+\varepsilon f$ is analytic and h satisfies a quite mild condition, of a geometric nature, which is called "steepness" and is essentially a weaker form of convexity (and that Nekhoroshev has proven to be, in some sense, generic). Moreover, a similar result also holds when h describes a system of uncoupled harmonic oscillators (which does not verify the steepness condition), under a suitable nonresonance condition [11].

It is of interest to compare Nekhoroshev's results with the other great theorem in Hamiltonian perturbation theory, namely KAM theorem. As it is well known (see for instance [2,6,37]), KAM theorem assures that (under suitable hypotheses, like nondegeneracy of h) the majority of invariant tori survives small perturbations, up to small deformations. More precisely, one can assures the existence of a set of initial conditions, which is of large measure but nowhere dense (its complement has measure which is small with ε and is open and dense), for which the motions are exactly almost periodic and, moreover, "close" (for all times) to the corresponding motions of the unperturbed system. However, KAM theorem fails to provide a stability result for all the motions, except in a very special case.

Indeed, when the unperturbed Hamiltonian h satisfies a "iso-energetic" nondegeneracy condition (see for instance [4]), then the majority of invariant tori on each energy submanifold survives the perturbation. For systems with two degrees of freedom, this fact implies the perpetual stability (boundedness) of all the motions, since any trajectory starting out from a non-surviving torus remains confined between two surviving tori around it. However, when the number of degrees of freedom exceeds two, this topological confinment does not exist any more, since (for $n>2$) the n-dimensional tori do not disconnect the $(2n-1)$-dimensional energy submanifolds. Consequently, there may exist unbounded motions which start from the gaps between the surviving tori and go far away in phase space. This phenomenon is called "Arnold diffusion", since Arnold provided an example of a system in which it actually occurs [3,6]; in this example, the velocity of the diffusion is exponentially small with the perturbation. In this context, Nekhoroshev theorem can be interpreted as establishing that Arnold's diffusion is a very slow phenomenon: indipendently of the complexity of the motions (homoclinic points can occur), the actions are "almost frozen" for times which are exponentially long in $1/\varepsilon$.

The ideas and the techniques used in the proof of Nekhoroshev theorem are of great relevance in perturbation theory, since they provide a very efficient overcoming of the "small denominators" difficulty. At the base of the proof there is a decomposition of the phase space into regions ("blocks"), characterized by definite local nonresonance conditions. The proof is then articulated in an "analytic" part, which deals with the construction, in each block, of a canonical transformation which gives the Hamiltonian, restricted to that block, a suitable "normal form" and in a "geometric" part, which has to provide a confinment of the actions in each block, assuring that the actions remain close to their initial values.

The purpose of the present thesis is to revisit the "analytic" part of Nekhoroshev theorem. As this part is based on (near to the identity) canonical transformations, an efficient canonical transformation theory is hopeful. The canonical transformation theory based on the so called "Lie transform" (or Lie series) techniques presents some advantages over the traditional one based on generating functions (the "mixed variable" method).

Thus, we first give in Chapter 1 a general and quite detailed introduction to Lie transforms. This presentation is also motivated by the fact that a comprehensive review of this subject is not available in the existing literature. Special attention is devoted to give rigorous existence results and to discuss a particular variant of the Lie technique -the so called "products of Lie series"- which will be subsequently used.

Chapter 2 deals with Nekhoroshev-like exponential estimates. We first provide an introduction to these results in the context of classical Hamiltonian perturbation theory and a brief outline of the whole proof of the theorem. We then present quite detailed proofs of the analyitc part of the theorem and of the results for the harmonic oscillators case (which does not require a "geometric" treatment at all). These proofs, which are deferred to the Appendixes, are performed by constructing the canonical transformations by means of products of Lie series and using a procedure inspired by a recent work of Galgani and Giorgilli [22].

CHAPTER 1

LIE SERIES AND LIE TRANSFORMS

1.1 INTRODUCTION

Traditionally, in the so called "mixed variable" or "Von Zeipel's" method, canonical transformations are described as a change of canonical coordinates $(p,q) \longrightarrow (P,Q)$ by means of a generating function which is given as a function of both some of the old and some of the new coordinates. For instance, given any function $S = S(P,q)$, the change of coordinates implicitly defined by the equations

$$(1) \qquad \begin{aligned} p &= \frac{\partial S}{\partial q}(P,q) \\ Q &= \frac{\partial S}{\partial P}(P,q) \end{aligned}$$

turns out to be canonical. Some variants of this scheme exist in which different sets of old and new coordinates are used, and indeed they have to be considered in order to describe any canonical transformation [4].

The case of interest in perturbation theory is that of near to the identity canonical transformations, namely that of a family of canonical transformations depending regularly on a parameter ε and reducing to the identity for $\varepsilon = 0$ (we will frequently not distinguish between the family and its single transformations, as no appreciable confusion arises). In this case the scheme of equations (1) turns out to be appropriate, with a generating function of the form $S(P,q,\varepsilon)=Pq + \widetilde{S}(P,q,\varepsilon)$.

Although any near to the identity canonical transformation can be described in this way and this method is in many aspects very efficient, it presents some inconveniences (for more details see for instance [15,16]): (i) inversions and substitutions are needed to explicitly obtain the transformation; (ii) the method does not lead to simple algorithms to obtain the transformation laws of functions or tensor fields; (iii) the machinery of generating functions is not canonically invariant and depends in an essential way on the coordinates; (iv) in all the analytic computations one is forced to use simultaneously two different coordinate system, with a cumbersome and awful notation.

Lie transform techniques constitute an alternative and very appealing possibility, as they do not suffer of these shortcomings. The basic idea is that of representing a near to the identity canonical transformation as (the map at some time of) the flow of a Hamiltonian vector field. Thus, in this context, a near to the identity canonical transformation is described as a diffeomorphism of phase space in purely geometrical terms, a unique coordinate system is used from the outset, no inversions are needed and, moreover, the transformation laws of functions are explicit and very simple. In particular, in the analytic case (which most directly interests the applications to perturbation theory) the canonical transformations and the transformations of functions that they induce are represented, via Taylor expansions, by power series in the parameter ε. These series are named Lie series; their coefficients are functions on phase space which can be evaluated with explicit algorithms.

The original idea of the method goes back to Lie [34] and has an obvious source in a Lie group context; quotations to some of the early references can be found in [20]; see also [24]. The systematic use of Lie transforms in connection with Hamiltonian perturbation theory began in the late sixthiees with Hori [25] and Deprit [15]. A number of contributions was then given to improve the algorithms, extend them to noncanonical transformations, propose new variants, etc. [12, 14-17,20,21,25-27,31,32,35,36,40]; for further references see the review papers [13,18,33]. Let us stress that most of this work was in the realm of formal power series; rigorous treatments, where the convergence problems are considered, are few [24,22,34]. Rigorous applications to Hamiltonian perturbation theory can be found in [7,8,9,22]. The geometrical setting for Lie transforms has been clearly discussed in [27].

In this chapter we give a general survey of this matter. We do not claim to be complete and try to focalize the attention on those topics which will be used in the following or which are not fully covered in the literature. Although in the sequel we will be concerned with the canonical case alone, we present here the general case as it does not introduce additional difficulties while permitting a simpler treatment.

We first give an elementary geometric presentation of the basic ideas of Lie transforms. Section 3 is devoted to the formal apparatus of Lie series. A rigorous treatment of the various cases (in the canonical case) is contained in the last section, where some technical devices to be used later are also included.

1.2 LIE TRANSFORMS

We consider a real (or complex), finite dimensional smooth manifold M and a one parameter family of smooth local diffeomorphisms on it, namely a smooth mapping $\psi : I \times U \longrightarrow M$, $(\varepsilon, z) \longmapsto \psi_\varepsilon(z)$ where I is a real (or complex) neighbourhood of zero and U an open subset of M, such that (i) for each $\varepsilon \in I$, ψ_ε is a smooth diffeomorphism of U onto its image (i.e. local) and (ii) ψ_0 is the identity on U. Usually, I will be a real interval $(-T, T)$ or a complex disk $B_T = \{t \in \mathbb{C} : |t| < T\}$ $(T>0)$.

The idea underlying the Lie method is that of representing the family of transformations $\psi : I \times U \rightarrow M$ as the local flow of a (possibly ε-dependent) vector field X defined by

(1) $$\frac{d\Psi_\varepsilon}{d\varepsilon} = X_\varepsilon \circ \Psi_\varepsilon \qquad\qquad (\varepsilon \in I)$$

We will discuss separately, for expositorial reasons, the two cases in which X depends or not on ε, which correspond to Lie's original idea and, respectively, to Deprit's generalization. Some variant of thic basic scheme are discussed in subsections C and D.

Let us stress that all the treatment is strictly local. In particular, equation (1) usually defines a local vector field on M, i.e. a vector field defined only on some open $U \subset M$.

1.1.A Simple Lie transforms

We first consider the case in which $\Psi : I \times U \to M$ is a local group, namely $\Psi_{\varepsilon_1+\varepsilon_2} = \Psi_{\varepsilon_1} \circ \Psi_{\varepsilon_2}$ for all $\varepsilon_1, \varepsilon_2, \varepsilon_1 + \varepsilon_2 \in I$. This case is obviously characterized by the fact that equation (1) defines a unique vector field X for all $\varepsilon \in I$, namely $X = \left(\frac{d\Psi_\varepsilon}{d\varepsilon}\right)_{\varepsilon=0}$; its local flow $\phi^X : (t,z) \mapsto \phi^X_t z$ is thus defined on $I \times U$ and there coincides with Ψ, in the sense that for each $\varepsilon \in I$ it is $\Psi_\varepsilon = \phi^X_\varepsilon$. Conversely, the local flow ϕ^X of any vector field X is defined on some open nonempty set $I \times U$ and there constitutes a local group of local diffeomorphism.

Let us stress that in this picture the parameter ε can be regarded as the time of the flow ϕ^X. It is also possible to represent each single transformation Ψ_ε as the time-one map of a flow by considering the family of vector fields $\{\varepsilon X\}_{\varepsilon \in I}$, each of which induces a local flow $\phi^{\varepsilon X}$ such that $\phi^{\varepsilon X}_{t=1} = \phi^X_\varepsilon = \psi_\varepsilon$.

Of central importance in transformation theory are the transformation laws of functions and tensor fields, which geometrically are given as pullbacks. For our purposes it will be sufficient to consider only the scalar case; the generalization to tensor fields is anyway trivial. By definition, the pullback of a function f under ϕ^X_ε is

(2) $\quad (\phi^X_\varepsilon)^* f = f \circ \phi^X_\varepsilon$

The family of transformations $f \mapsto (\phi^X_\varepsilon)^* f$ (for $\varepsilon \in I$) will be called the (simple) Lie transform of f generated by the vector field X.

Coordinate description. At the chart level, Lie transforms can also be described as the change of functions under a change of coordinates. For definitness let us assume that $\bigcup_{\varepsilon \in I} \phi^X_\varepsilon(U)$ is contained in the domain of a single chart with coordinates $z = (z^1, \ldots, z^n)$. Let us denote X_z the local representative in this chart of the vector field X and $\mathcal{F} = f \circ z^{-1}$ that of a function f. Since the flow of X is represented by $z \circ \phi^X_\varepsilon \circ z^{-1} = \phi^{X_z}_\varepsilon$, i.e. by the flow of X_z, the function $(\phi^X_\varepsilon)^* f$ has the representative $F_\varepsilon = (\phi^{X_z}_\varepsilon)^* \mathcal{F}$. So, the Lie transform is represented in this coordinate system by

(3) $\quad \mathcal{F} \mapsto F_\varepsilon = (\phi^{X_z}_\varepsilon)^* \mathcal{F}$

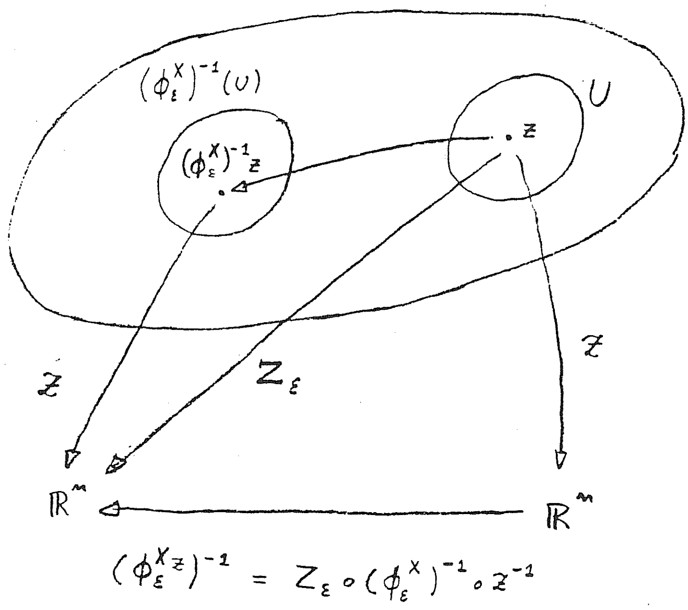

figure 1

On the other hand, $Z_\varepsilon^i = [(\phi_\varepsilon^X)^{-1}]^* \mathfrak{z}^i$ ($i=1,\dots,n$) are new coordinates on U since the transition mapping $Z_\varepsilon \circ \mathfrak{z}^{-1} = \mathfrak{z} \circ (\phi_\varepsilon^X)^{-1} \circ \mathfrak{z}^{-1} = (\phi_\varepsilon^{X_{\mathfrak{z}}})^{-1}$ is a local diffeomorphism onto $\mathfrak{z}(U)$ (figure 1). In the Z_ε-coordinates the local representative of f is precisely F_ε, as it results from $f = \mathcal{F} \circ \mathfrak{z} = F_\varepsilon \circ Z_\varepsilon$. Thus (3) can also be interpreted as the transformation law of the representative of f under the transformation of coordinates $\mathfrak{z} \mapsto Z_\varepsilon = [(\phi_\varepsilon^X)^{-1}]^* \mathfrak{z}$.

Canonical Lie transforms Canonical transformations are naturally recovered in this setting: if M is a symplectic manifold all that one needs is to consider (locally) Hamiltonian vector fields (for these notions see for instance [1]). Indeed, if X is (locally) Hamiltonian with a Hamilton function $\mathcal{X}$, then for each $\varepsilon \in I$ the diffeomorphism ϕ_ε^X is canonical. Conversely, if the family $\{\psi_\varepsilon\}_{\varepsilon \in I}$ is canonical, then the vector field X defined by (1) is Hamiltonian and a Hamilton function $\mathcal{X}$ for it exists.

In this case we will usually write $\phi^{\mathcal{X}}$ for ϕ^{X} and we will call $\mathcal{X}$ the generating function of the canonical Lie transform.

1.2.B. Deprit's Lie transforms

In geometrical terms, Deprit's generalization of the previous case was simply that of considering the case of ε-dependent vector fields X defined by (1), so to describe in this way any family of near to the identity canonical transformations.

Let us recall that a time dependent, or non-conservative, (local) vector field is a map $X:(t,z) \mapsto X_t(z)$ from some subset of $\mathbb{R}\times M$ (or $\mathbb{C}\times M$) into TM such that, for each t, X_t is a (local) vector field on M. We denote $\phi^X : I\times I\times U \to M$, $(t_1, t_2, z) \mapsto \phi^X_{t_1,t_2} z$ its (local) "nonsteady" flow defined by $\left(\frac{d}{dt}\phi^X_{t_1,t}\right)_{t=t_2} = X_{t_2} \circ \phi^X_{t_1,t_2}$, $\phi^X_{t,t} = \mathbb{1}$.

Consider now the mappings $\phi^X_{0,\cdot} : I\times U \to M$, $(t,z) \longmapsto \phi^X_{0,t} z$. For $t\in I$ they constitute a family of local diffeomorphisms. Conversely, given a family $\{\psi_\varepsilon\}_{\varepsilon\in I}$ of local diffeomorphisms it results $\psi_\varepsilon = \phi^X_{0,\varepsilon}$ for all $\varepsilon\in I$ if X is defined as in (1).

Transformation of functions is again given by the pullback under the flow, i.e. $f \mapsto (\phi^X_{0,\varepsilon})^* f = f\circ\phi^X_{0,\varepsilon}$, and this map will be called the Deprit's Lie transform of f generated by X. Obviously it reduces to a simple Lie transform when X is conservative.

The canonical case is recovered as before. If, for each $\varepsilon \in I$, X_ε is a Hamiltonian vector field, then the mapping ϕ^X_ε is canonical. Of course an ε-dependent Hamilton function $\mathcal{H}_\varepsilon$ corresponds to X_ε.

Remarks (i) The inverse of $\phi^X_{0,\varepsilon}$ is $(\phi^X_{0,\varepsilon})^{-1} = \phi^X_{0,\varepsilon}$; when X is Hamiltonian, it is canonical too. Thus given a non-conservative vector field X one can consider, besides the Deprit's transform $f \mapsto (\phi^X_{0,\varepsilon})^* f$ also its inverse $f \mapsto [(\phi^X_{0,\varepsilon})^{-1}]^* f = f \circ (\phi^X_{0,\varepsilon})^{-1}$. Actually, as it will be discussed in the next section, the use of the inverse transformations presents some advantages.

(ii) It often happens that the function f to be transformed depends also on the parameter ε. The transformed of such a function under $\phi^X_{0,\varepsilon}$ is, at each ε, the function $z \mapsto f(\phi^X_{0,\varepsilon} z, \varepsilon)$. Writing $f: (z,\varepsilon) \longmapsto f_\varepsilon(z)$, we can then write the Deprit's transform of f, at a given ε, as $(\phi^X_{0,\varepsilon})^* f_\varepsilon = f \circ \phi^X_{0,\varepsilon}$: this can be interpreted by saying that one acts with the flow ϕ^X of X between the "times" $t=0$ and $t=\varepsilon$ on the "final" form f_ε of the function; to stress this fact one can also note that, obviously, $(\phi^X_{0,\varepsilon})^* f_\varepsilon = (f_\varepsilon \circ \phi^X_{0,t})_{t=\varepsilon}$. In particular, it follows from this that, for any function g, one has $(\phi^X_{0,\varepsilon})^* (\varepsilon g) = \varepsilon (\phi^X_{0,\varepsilon})^* g$ and then that $(\phi^X_{0,\varepsilon})^*$ acts linearly on power series in ε:

$$(4) \qquad (\phi^X_{0,\varepsilon})^* \sum_{s \geq 0} \varepsilon^s g_s = \sum_{s \geq 0} \varepsilon^s (\phi^X_{0,\varepsilon})^* g_s$$

1.2.C. Hori's Lie transforms

Hori's generalization of Lie transforms, which chronologically preceded that of Deprit, is based on a slightly different idea. Here an ε-dependent vector field Y_ε ($\varepsilon \in I$) is given, but it is regarded as a family $\{Y_\varepsilon\}_{\varepsilon \in I}$ of conservative vector fields; i.e. for each fixed $\varepsilon \in I$ the flow $\phi^{Y_\varepsilon}: (t, z) \mapsto \phi_t^{Y_\varepsilon} z$ of Y_ε is considered. Then for each $\varepsilon \in I$ one considers the map at time $t = \varepsilon$ of this flow, i.e. $\phi_{t=\varepsilon}^{Y_\varepsilon}$. By taking all these mappings one obtains a family of near to the identity (local) diffeomorphisms $\{\phi_{t=\varepsilon}^{Y_\varepsilon}\}_{\varepsilon \in I}$. The transformation $f \mapsto (\phi_{t=\varepsilon}^{Y_\varepsilon})^* f$ is then said to be the Hori's Lie transform generated by Y.

It turns out that any near to the identity transformation (for sufficiently small ε) can be represented in this way since, as shown in [28], given a non-conservative vector field X there exists another one Y such that for each ε it is $\phi_{0,\varepsilon}^X = \phi_{t=\varepsilon}^{Y_\varepsilon}$, and conversely. Explicit relations between the two generating vector fields X and Y have been obtained in the analytic, canonical case [14,35,28]. Let us note that Hori's generator Y can also be used to represent, for each ε, $\phi_{0,\varepsilon}^X$ as the time-one map of a flow, since it is $\phi_{0,\varepsilon}^X = \phi_{t=1}^{\varepsilon Y_\varepsilon}$.

Hori's Lie transform is not very used in applications since, in the analytic case, it does not give rise to a general algorithm to evaluate the transformation of functions $f \mapsto (\phi_{t=\varepsilon}^{Y_\varepsilon})^* f$ (see however [21]); accordingly, we will not consider it any more in the following.

1.2.D. Composition of simple Lie transforms

This is a different possibility, which is completely natural in perturbation theory since it corresponds to the intuitive idea of performing the normalization procedure step by step. The idea, first due to Dragt and Finn [17], is that of considering a collection of vector fields $X_1,\dots,X_r$ and then compose the maps at times $t_1=\varepsilon$, $t_2=\varepsilon^2,\dots,$ $t_r=\varepsilon^r$ of their local flows to obtain a mapping $(\varepsilon,z)\longmapsto \phi_\varepsilon^{(X_1,\dots,X_r)}(z) \equiv \phi_\varepsilon^{X_1}\circ\phi_{\varepsilon^2}^{X_2}\circ\dots\circ\phi_{\varepsilon^r}^{X_r}(z)$.

To be more precise, let us consider a finite number of vector fields $X_1,\dots,X_r$ whose local flows are defined on a common set $I\times U$. By the continuity of the flows, for each $s=2,\dots,r$ there exists a nonemty open set $I_s\times U_s$, with $U_r\subset\dots\subset U_2\subset U_1=U$, such that $\bigcup_{t_s\in I_s}\phi_{t_s}^{X_s}(U_s)\subset U_{s-1}$. Thus, with any $I_1\subset I$, the mapping $\phi_{t_1,\dots,t_r}^{(X_1,\dots,X_r)}: I_1\times\dots\times I_r\times U_r\longrightarrow M$, $(t_1,\dots,t_r,z)\longmapsto \phi_{t_1,\dots,t_r}^{(X_1,\dots,X_r)}(z) = \phi_{t_1}^{X_1}\circ\dots\circ\phi_{t_r}^{X_r}(z)$ is well defined; moreover, for each $(t_1,\dots,t_r)\in I_1\times\dots\times I_r$, $\phi_{t_1,\dots,t_r}^{(X_1,\dots,X_r)}$ is a diffeomorphism of U_r onto its image, its inverse being $\phi_{-t_r,\dots,-t_1}^{(X_r,\dots,X_1)} = \phi_{-t_r}^{X_r}\circ\dots\circ\phi_{-t_1}^{X_1}$. The transformation of functions $f\longmapsto \phi_{t_1,\dots,t_r}^{(X_1,\dots,X_r)*}f$ is then obviously given by the composition (product) of the corresponding simple Lie transforms, but in the reversed order: $\phi_{t_1,\dots,t_r}^{(X_1,\dots,X_r)*}f = f\circ(\phi_{t_1}^{X_1}\circ\dots\circ\phi_{t_r}^{X_r}) = [(\phi_{t_r}^{X_r})^*\dots(\phi_{t_1}^{X_1})^*]\,f$.

The case of interest is that in which one considers the restriction of this mapping at times t_s which are different powers of a single parameter ε: $t_s=\varepsilon^s$ for $s=1,\dots,r$. Thus, with

$$(5) \qquad \phi_\varepsilon^{(X_1,\dots,X_r)} \equiv \phi^{(X_1,\dots,X_r)}_{t_1=\varepsilon,\dots,t_r=\varepsilon^r} \equiv \phi_\varepsilon^{X_1}\circ\dots\circ\phi_{\varepsilon^r}^{X_r}$$

one defines a one-parameter family of local diffeomorphisms $\phi_\varepsilon^{(X_1,\dots,X_r)}: J\times V \to M$, $(\varepsilon,z)\longmapsto \phi_\varepsilon^{(X_1,\dots,X_r)}(z)$ (for some suitable $J\times V$). Note that these diffeomorphisms could also be regarded as compositions of time-one maps of different flows, since one trivially has $\phi^{(X_1,\dots,X_r)}_{t_1=\varepsilon,\dots,t_r=\varepsilon^r} = \phi^{(\varepsilon X_1,\dots,\varepsilon^r X_r)}_{t_1=1,\dots,t_r=1}$.

The transformation of a function f under every transformation (5) is given by

$$(6) \qquad \phi_\varepsilon^{(X_1,\dots,X_n)*} f = (\phi_{\varepsilon^r}^{X_r})^*\cdot\dots\cdot(\phi_\varepsilon^{X_1})^* f$$

and we will refer to it as to the composition of Lie transforms generated by X ,...,X .

Remarks (i) The canonical case is recovered by using Hamiltonian vector fields. If $\mathcal{X}_1,\dots,\mathcal{X}_r$ are their Hamilton functions, we will write $\phi_\varepsilon^{(\mathcal{X}_1,\dots,\mathcal{X}_r)}$.

(ii) One could ask for a (nonconservative) vector field Y such that $\phi^Y_{0,\varepsilon} = \phi_\varepsilon^{(X_1,\dots,X_r)}$ for each ε. This problem has been investigated, in the realm of formal power series, in [18].

(iii) Not all the near to the identity transformations can be represented as a composition of a finite number of flows (while they could be so approximated [35]). The composition of an infinite number of flows could be treated along the same lines, as no convergence problem arise (see the proposition 1.2 in the next section); the only point which would require some care is the invertibility of such a mapping.

1.3 LIE SERIES

Lie transform techniques become a powerful instrument in the analytic case, because the transformed functions can be represented, via Taylor expansions, as power series in the "small" parameter ε. A number of algorithms have been developed for the evaluation of the coefficients of these "Lie series". This section is devoted to a presentation of this formal apparatus, which has an obvious relevance for the applications to perturbation theory. If a geometric description is desired, analytic manifolds (real or complex) have to be used; since there is no advantage in this generality one can simply think to work on some open subset of R^n or C^n.

1.3.A. Simple Lie series

We first discuss the trivial case of simple Lie transforms, as it serves as introduction to the others. When X is analytic, its local flow ϕ^X is analytic on some open nonemty set $I \times U$; for convenience, let us take for I a real interval $(-T,T)$ or a complex disk $B_T = \{t \in C : |t| < T\}$. If f is any analytic function (defined at least on $\bigcup_{t \in I} \phi_t^X(U)$), the function $f \circ \phi^X : I \times U \to R$ (or C) is analytic too and can then be expanded, for each $z \in U$, in its Taylor series around $\varepsilon = 0$:

$$(1) \qquad (\phi^X_\varepsilon)^* f = \sum_{s=0}^{\infty} \frac{\varepsilon^s}{s!} \frac{d^s}{dt^s} (f \circ \phi^X_t)_{t=0}$$

The convergence of this series is a trivial matter in the complex case, since it is uniformly convergent in $B_T \times U$; in the real case some care has to be taken on the possible presence of singularities of $\varepsilon \mapsto f \circ \phi^X_\varepsilon$ outside the real axis.

The usefulness of the Taylor expansion is due to the fact that to compute it one does not need to know in advance the flow ϕ^X, because of the basic relation

$$(2) \qquad \frac{d^s}{dt^s} (f \circ \phi^X_t) = (L^s_X f) \circ \phi^X_t$$

where $L^0_X = \mathbb{1}$, $L^{s+1}_X = L_X L^s_X$ and L_X is the Lie derivative associated with X. Recall that, by definition, $L_X f(z) = df(z) \cdot X(z)$ or, in coordinates, $L_X f = \sum_{i=1}^{n} X^i \frac{\partial f}{\partial z^i}$ (for the details see [1]). Thus, the Taylor expansion (1) can be written

$$(3) \qquad (\phi^X_\varepsilon)^* f = \sum_{s=0}^{\infty} \frac{\varepsilon^s}{s!} L^s_X f$$

which is called the Lie series for f generated by the vector field X; clearly, it can be formally written $\exp(\varepsilon L_X) f$.

In the canonical case it is also possible to express the Lie series in terms of a Hamilton function $\mathcal{X}$ of X. Let us introduce the differential operator $L_{\mathcal{X}} = \{\cdot, \mathcal{X}\}$, where $\{\ ,\ \}$ are Poisson brackets (in canonical coordinates (p,q): $\{f, g\} = \sum_{i=1}^{n} \left[\frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q_i} - \frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i} \right]$). Now, if X has the Hamilton function

$\mathcal{X}$, it is $L_X = L_{\mathcal{X}}$, so that the Lie series $\exp(\varepsilon L_X)$ can also be written $\exp(\varepsilon L_{\mathcal{X}}) = \sum_{s\geq 0} \frac{\varepsilon^s}{s!} L^s_{\mathcal{X}}$.

1.3.B. Deprit's Lie series

Taylor series can again be used to represent the pullback of analytic functions under an analytic nonsteady flow but, at difference with the previous case, "time" derivatives are not simply given by Lie derivatives and some more work is needed to obtain useful forms for the series. Indeed, while it is

$$(4) \qquad \frac{d}{dt}\left(f \circ \phi^X_{0,t}\right)_{t=\varepsilon} = \left(L_{X_\varepsilon} f\right) \circ \phi^X_{0,\varepsilon}$$

for higher derivatives also the ε-dependence of X has to be taken into account, so that (for all $s\geq 0$):

$$(5) \qquad \frac{d^s}{dt^s}\left(f \circ \phi^X_{0,t}\right)_{t=\varepsilon} = \left[\left(\frac{\partial}{\partial t} + L_{X_t}\right)^s f\right]_{t=\varepsilon} \circ \phi^X_{0,\varepsilon}$$

These formulas are easily established by using the definition of the Lie derivative L_{X_ε} (at each ε); in (5) it is $\left(\frac{\partial}{\partial t} + L_{X_t}\right)^0 = \mathbb{1}$, $\left(\frac{\partial}{\partial t} + L_{X_t}\right)^{s+1} = \left(\frac{\partial}{\partial t} + L_{X_t}\right)^s \left(\frac{\partial}{\partial t} + L_{X_t}\right)$ and the usual differentiation rules apply. In the following we will write

$$(6) \qquad \mathcal{F}^{(s)}[f] \equiv \frac{d^s}{dt^s}\left(f \circ \phi^X_{0,t}\right)_{t=0} = \left[\left(\frac{\partial}{\partial t} + L_{X_t}\right)^s f\right]_{t=0}$$

so that (for sufficiently small ε) one has

(7) $$(\phi^X_{0,\varepsilon})^* f = \sum_{s=0}^{\infty} \frac{\varepsilon^s}{s!} \mathcal{F}^{(s)}[f]$$

which will be called the Deprit's Lie series for f generated by X (we will often write $f^{(s)}$ for $\mathcal{F}^{(s)}[f]$).

The success of this scheme is due to the fact that it is possible to give expressions for the $\mathcal{F}^{(s)}$ from which all the ε-derivatives and all the ε-dependence of X_ε have been eliminated. The key for this is in the use of the Taylor expansion of X_ε around $\varepsilon=0$, say (with $X_1=X_{\varepsilon=0}$, $X_{s+1} = \left(\frac{\partial^s X_\varepsilon}{\partial \varepsilon^s}\right)_{\varepsilon=0}$)

(8) $$X_\varepsilon = \sum_{s=0}^{\infty} \frac{\varepsilon^s}{s!} X_{s+1}$$

In this way, Deprit [15] succeeded in obtaining an algorithm to recursively evaluate all the $\mathcal{F}^{(s)}$ by operating with certain combination of the L_{X_j} ($j=1,2,..$). Some variants of this algorithm have been proposed in [26,31,32]. We will present here only one of these algorithms and discuss some related topics.

Remarks (i) The use of the series (8) emerges naturally when considering the "mixed" derivatives in (6). Noting that $\frac{\partial}{\partial\varepsilon} L_{X_\varepsilon} = L_{\frac{\partial X_\varepsilon}{\partial \varepsilon}}$ and using Leibnitz rule (with some care to the noncommutativity of the vector fields $\frac{\partial^s X_\varepsilon}{\partial \varepsilon^s}$ with different s) one sees that, for instance, $\frac{\partial^m}{\partial\varepsilon^m} L^p_{X_\varepsilon} f = \sum_{k_1,\dots,k_p=0,\dots,m} c_{k_1,\dots,k_p} L_{\frac{\partial^{k_1} X_\varepsilon}{\partial\varepsilon^{k_1}}} \cdot \cdots L_{\frac{\partial^{k_p} X_\varepsilon}{\partial\varepsilon^{k_p}}}$, where the $c_{k_1,\dots,k_p}$ are some products of binomial coefficients. All the derivatives in (6), being evaluated at $\varepsilon=0$, reduce then to combination of the L_{X_j}.

(ii) It is of course possible to write Deprit's series in exponential form passing to the extended space R^{n+1} (or C^{n+1}). With $\hat{X} = X + \frac{\partial}{\partial t} : (t,z) \mapsto X_t(z) + \frac{\partial}{\partial t}$ it is $L_{\hat{X}} = L_X + \frac{\partial}{\partial t}$ and then $\sum_{s=0}^{\infty} \frac{\varepsilon^s}{s!} \left[\left(\frac{\partial}{\partial \varepsilon} + L_{X_\varepsilon} \right)^s f \right]_{\varepsilon=0} = \left[\exp(\varepsilon L_{\hat{X}}) f \right](z,0)$.

(iii) The canonical case can be treated as in the case of simple Lie series. If $\mathcal{K}_\varepsilon = \sum_{s=0}^{\infty} \frac{\varepsilon^s}{s!} \mathcal{K}_{s+1}$ is a Hamilton function for X_ε, then each $\mathcal{K}_s$ is a Hamilton function for the corresponding X_s and the Lie derivatives L_{X_s} can also be regarded as the Poisson brackets $L_{\mathcal{K}_s}$.

Deprit's inverse series. Besides the transformation $f \mapsto (\phi^X_{0,\varepsilon})^* f$ one can also consider the transformation of functions under the inverse of the diffeomorphism $\phi^X_{0,\varepsilon}$, namely $f \mapsto [(\phi^X_{0,\varepsilon})^{-1}]^* f$, whose Taylor series expansion will be written

$$\text{(9)} \qquad \left[(\phi^X_{0,\varepsilon})^{-1} \right]^* f = \sum_{s=0}^{\infty} \frac{\varepsilon^s}{s!} \mathcal{T}_{(s)}[f]$$

with

$$\text{(10)} \qquad \mathcal{T}_{(s)}[f] \equiv \frac{d^s}{d\varepsilon^s} \left[f \circ (\phi^X_{0,\varepsilon})^{-1} \right]_{\varepsilon=0}$$

(usually written $\mathcal{T}_{(s)}$). As first noted by Kamel [30] and Henrard [26], the functions $\mathcal{T}_{(s)}$ can be computed with a recursive algorithm which is more efficient than those for the $\mathcal{T}^{(s)}$. Thus, both in rigorous as well in numerical applications it could be convenient to describe near to the identity transformations with the inverse of the nonsteady flow $\phi^X_{0,\varepsilon}$. The recursive "algorithm of the inverse" has been obtained by inverting in a purely algebraic way the direct one. It is also possible, and simpler, to obtain it by directly working out equation (10), as we will do in the following; the basic relation for this, proved later, is

$$(11) \qquad \frac{d}{dt}\left[f\circ(\phi^X_{0,\varepsilon})^{-1}\right]_{t=\varepsilon} = - L_{X_\varepsilon}\left[f\circ(\phi^X_{0,\varepsilon})^{-1}\right]$$

Transformation of ε-dependent functions. The recursive algorithms for Deprit's series are usually obtained in the more general case in which the function to be transformed depends also on ε. The employed procedure is that of regarding the Deprit's series as the Taylor series of the function $\varepsilon \mapsto f(\phi^X_{0,\varepsilon} z, \varepsilon) = [(\phi^X_{0,\varepsilon})^* f_\varepsilon](z)$ (where $f_\varepsilon(z) = f(z, \varepsilon)$), namely

$$(\phi^X_{0,\varepsilon})^* f_\varepsilon = \sum_{s=0}^{\infty} \frac{\varepsilon^s}{s!} \mathcal{T}^{(s)}[f]$$

where

$$\mathcal{T}^{(s)}[f] = \left[\frac{d^s}{d\varepsilon^s}(f_\varepsilon\circ\phi^X_{0,\varepsilon})\right]_{\varepsilon=0} = \frac{d^s}{d\varepsilon^s}\left[f(\phi^X_{0,\varepsilon} z, \varepsilon)\right]_{\varepsilon=0}$$

These ε-derivatives can again be expressed as in (5) or (6), but now also ε-derivatives of f_ε have to be performed; for instance, $[(\frac{\partial}{\partial\varepsilon} + L_{X_\varepsilon}) f_\varepsilon]_{\varepsilon=0} = (\frac{\partial f_\varepsilon}{\partial\varepsilon})_{\varepsilon=0} + L_{X_{\varepsilon=0}} f_{\varepsilon=0}$. The evaluation of the ε-derivatives of f_ε is performed by means of the Taylor series in ε of this function, say $f_\varepsilon = \sum_{s=0}^{\infty} \frac{\varepsilon^s}{s!} f_s$; the final result is then a recursion for the $\mathcal{T}^{(s)}[f_\varepsilon]$ written in terms of the Lie derivatives L_{X_j} and of the functions f_j (it can be obtained from equations (12) and (15) below). For the details see, for instance, [26,29].

However, there is an alternative, and simpler, procedure which does not require ε-derivatives of f_ε at all. Indeed, by using equation (1.2.4), which was

$$(\phi^X_{0,\varepsilon})^* \sum_{s=0}^{\infty} \frac{\varepsilon^s}{s!} f_s = \sum_{s=0}^{\infty} \frac{\varepsilon^s}{s!} (\phi^X_{0,\varepsilon})^* f_s$$

and the series expansion $f_\varepsilon = \sum_{s=0}^{\infty} \frac{\varepsilon^s}{s!} f_s$, one immediately sees that

$$(12) \qquad \mathcal{F}^{(s)}[f_\varepsilon] = \sum_{j=0}^{s} \binom{s}{j} \mathcal{F}^{(j)}[f_{s-j}]$$

where $\binom{s}{j} = \frac{s!}{(s-j)!\,j!}$ are the binomial coefficients. Thus, known any algorithm to generate the different power terms of the Deprit's Lie series of ε-independent functions, one immediately recovers the general case. Let us stress that (12) could be equally well obtained by noting that

$$(13) \qquad (\phi^X_{0,\varepsilon})^* f = (f_\varepsilon \circ \phi^X_{0,t})_{t=\varepsilon} = \sum_{s=0}^{\infty} \frac{\varepsilon^s}{s!} \left(\frac{d^s}{dt^s} f_\varepsilon \circ \phi^X_{0,t} \right)_{t=0}$$

and substituting here for f_ε its power series in ε; this procedure obviously coincides with the former one because of the uniqueness of the Taylor expansion. Similarly, for the inverse transformation one gets

$$(14) \qquad \mathcal{F}_{(s)}[f_\varepsilon] = \sum_{j=0}^{s} \binom{s}{j} \mathcal{F}_{(j)}[f_{s-j}]$$

For this reason we present here only the algorithms for ε-independent functions.

Direct Algorithm (Kamel and Henrard). Let $X_\varepsilon = \sum_{s=0}^{\infty} \frac{\varepsilon^s}{s!} X_{s+1}$ and, for all $s \geqslant 1$, $L_{X_s} = L_s$. Define recursively the differential operators

$$
(15) \qquad \begin{aligned} M_0 &= 1 \\ M_s &= \sum_{j=1}^{s} \binom{s-1}{j-1} M_{s-j} L_j \qquad\qquad (s \geqslant 1) \end{aligned}
$$

Then, for any function f and any $s \geqslant 0$, it is (with the notation of (6)) $\mathcal{F}^{(s)}[f] = M_s f$.

Proof. For convenience, let us prove the equivalent form $M_{s+1} = \sum_{j=0}^{s} \binom{s}{j} M_{s-j} L_{j+1}$ for $s \geqslant 0$. Let us introduce the differential operator $D_\varepsilon = \frac{\partial}{\partial\varepsilon} + L_{X_\varepsilon}$, so that for any ($\varepsilon$-independent) function g and all $s \geqslant 0$ it is $(\frac{\partial}{\partial\varepsilon} + L_{X_\varepsilon})^s g = [D_\varepsilon^s g]_{\varepsilon=0} \equiv (D_0)^s g$. Clearly $(D_0)^0 = 1 = M_0$. Proceeding by induction we show that, if it is also $(D_0)^1 = M_1, \ldots, (D_0)^s = M_s$, then $(D_0)^{s+1} = M_{s+1}$. For each $s \geq 0$ it is

$$
(D_0)^{s+1} f \equiv \left[D_\varepsilon^{s+1} f\right]_{\varepsilon=0} = \left[D_\varepsilon^s (L_{X_\varepsilon} f)\right]_{\varepsilon=0} = \left\{ D_\varepsilon^s \left[\sum_{p=0}^{\infty} \frac{\varepsilon^p}{p!} L_{p+1} f \right] \right\}_{\varepsilon=0}
$$

and using the linearity of D_ε and then Leibnitz rule to compute $D_\varepsilon^s (\varepsilon^p L_{p+1} f)$ this gives

$$
(D_0)^{s+1} f = \sum_{p=0}^{\infty} \sum_{j=0}^{s} \binom{s}{j} \left[D_\varepsilon^j \left(\frac{\varepsilon^p}{p!} \right) \right]_{\varepsilon=0} \left[D_\varepsilon^{s-j} (L_{p+1} f) \right]_{\varepsilon=0}
$$

But $L_{X_\varepsilon}(\varepsilon^p) = 0$, so that $D_\varepsilon^j (\varepsilon^p) = \frac{\partial^p}{\partial \varepsilon^p} (\varepsilon^p)$; thus

$$
(D_0)^{s+1} f = \sum_{j=0}^{s} \binom{s}{j} \left[D_\varepsilon^{s-j} (L_{j+1} f) \right]_{\varepsilon=0}
$$

But for the induction hypothesis it is $[D_\varepsilon^{s-j}(L_{j+1}f)]_{\varepsilon=0} = M_{s-j}(L_{j+1}f)$ for all $0 \le j \le s$ so that $(D_0)^{s+1} f = M_{s+1} f$. ■

To give an idea of how the operators M_s look we give the first few of them:

$$M_1 = L_1$$
$$M_2 = L_1^2 + L_2$$
$$M_3 = L_1^3 + L_2 L_1 + 2L_1 L_2 + L_3$$
$$M_4 = L_1^4 + L_2 L_1^2 + 2L_1 L_2 L_1 + L_3 L_1 + 3L_1^2 L_2 + 3L_2^2 + 3L_1 L_3 + L_4$$

Note that, according to (15), the number of addends in M_s is 2^{s-1}.

This algorithm has the nice feature of giving a simple formula to simbolically generate the operators M_s and thus to evaluate the functions $\mathcal{F}^{(s)} = M_s f$ (in this respect, Deprit's original algorithm is much more cumbersome). On the other hand it does not give a recursion between the $\mathcal{F}^{(s)}$ alone: because of the appearance of the L_j on the right of the M_{s-j} in (15) it is not possible to obtain each $\mathcal{F}^{(s)}$ by directly operating only on $\mathcal{F}^{(0)}, \mathcal{F}^{(1)}, \ldots, \mathcal{F}^{(s-1)}$ and then, in recursive computations, one has to re-evaluate at each step a great number of Lie de, ivatives. This indeed a common characteristic of all the known direct algorithms.

The advantage of the algorithm of the inverse is precisely that it does not suffer of this defect since the $\mathcal{F}_{(s)}$ are recursively defined by directly acting with the L_j on $\mathcal{F}_{(0)}, \ldots, \mathcal{F}_{(s-1)}$:

Algorithm of the inverse (Kamel and Henrard) Let $X_\varepsilon = \sum_{s\geq 0} \frac{\varepsilon^s}{s!} X_{s+1}$, write $L_s = L_{X_s}$ and, for any function f, define $\mathcal{F}_{(s)}$ as in (10). Then

$$(16) \qquad \begin{aligned} \mathcal{F}_{(0)} &= f \\ \mathcal{F}_{(s)} &= -\sum_{j=1}^{s} \binom{s-1}{j-1} L_j \mathcal{F}_{(s-j)} \end{aligned}$$

Proof Clearly $\mathcal{F}_{(0)} = f$. For $s \geq 1$ we prove (16) with a direct computation based on equation (11), which is (with $\phi_{0,\varepsilon} = \phi^X_{0,\varepsilon}$)

$$(11) \qquad \frac{d}{d\varepsilon}\left(f \circ \phi^{-1}_{0,\varepsilon}\right) = -L_{X_\varepsilon}\left(f \circ \phi^{-1}_{0,\varepsilon}\right)$$

and is proved below. For each $s \geq 1$ we can write, using (11):

$$\frac{d^{s+1}}{d\varepsilon^{s+1}}\left(f \circ \phi^{-1}_{0,\varepsilon}\right) = -\frac{d^s}{d\varepsilon^s}\left[L_{X_\varepsilon}\left(f \circ \phi^{-1}_{0,\varepsilon}\right)\right]$$

Note now that $\frac{d}{d\varepsilon}\left[L_{X_\varepsilon}(f\circ\phi^{-1}_{0,\varepsilon})\right] = L_{\frac{\partial X_\varepsilon}{\partial\varepsilon}}(f\circ\phi^{-1}_{0,\varepsilon}) + L_{X_\varepsilon}\frac{d}{d\varepsilon}(f\circ\phi^{-1}_{0,\varepsilon})$ so that, using Leibnitz rule, we get for all $s\geq 0$:

$$\frac{d^s}{d\varepsilon^s}\left[L_{X_\varepsilon}(f\circ\phi^{-1}_{0,\varepsilon})\right] = \sum_{j=0}^{s}\binom{s}{j} L_{\frac{\partial^j X_\varepsilon}{\partial\varepsilon^j}}\left[\frac{d^{s-j}}{d\varepsilon^{s-j}}(f\circ\phi^{-1}_{0,\varepsilon})\right]$$

So

$$\mathcal{F}_{(s+1)} \equiv \left[\frac{d^{s+1}}{d\varepsilon^{s+1}} f\circ\phi^{-1}_{0,\varepsilon}\right]_{\varepsilon=0} = -\sum_{j=0}^{s}\binom{s}{j} L_{\left(\frac{\partial^j X_\varepsilon}{\partial\varepsilon^j}\right)_{\varepsilon=0}}\left[\frac{d^{s-j}}{d\varepsilon^{s-j}}(f\circ\phi^{-1}_{0,\varepsilon})\right]_{\varepsilon=0}$$

or, by the definition of X_{j+1} and $\mathcal{F}_{(s-j)}$:

$$\mathcal{F}_{(s+1)} = -\sum_{j=0}^{s}\binom{s}{j} L_{j+1}\mathcal{F}_{(s-j)}$$

which gives (16) after a change of the summation index.

We now proof equation (11). From

$$0 = \frac{d}{d\varepsilon}[\phi_{0,\varepsilon} \circ \phi^{-1}_{0,\varepsilon}(z)] = \frac{\partial \phi_{0,\varepsilon}}{\partial \varepsilon} \circ \phi^{-1}_{0,\varepsilon}(z) + [d\phi_{0,\varepsilon}(\phi^{-1}_{0,\varepsilon} z)] \cdot \frac{\partial \phi^{-1}_{0,\varepsilon}}{\partial \varepsilon}(z)$$

where $d\phi_{0,\varepsilon}(\phi^{-1}_{0,\varepsilon} z)$ is the differential of $\phi_{0,\varepsilon}$ at the point $\phi^{-1}_{0,\varepsilon} z$, it follows that

$$\frac{\partial \phi^{-1}_{0,\varepsilon}}{\partial \varepsilon}(z) = -\left[d\phi_{0,\varepsilon}(\phi^{-1}_{0,\varepsilon} z)\right]^{-1} \left[\frac{\partial \phi_{0,\varepsilon}}{\partial \varepsilon} \circ \phi^{-1}_{0,\varepsilon}(z)\right].$$

namely

$$\frac{\partial \phi^{-1}_{0,\varepsilon}}{\partial \varepsilon} = -\left[(d\phi_{0,\varepsilon})^{-1} \frac{\partial \phi_{0,\varepsilon}}{\partial \varepsilon}\right] \circ \phi^{-1}_{0,\varepsilon}$$

Since it is $\frac{\partial \phi_{0,\varepsilon}}{\partial \varepsilon} = X_\varepsilon \circ \phi_{0,\varepsilon}$ and $(d\phi_{0,\varepsilon})^{-1} \cdot X_\varepsilon \circ \phi_{0,\varepsilon} = \phi^*_{0,\varepsilon} X_\varepsilon$ is the pullback of X_ε under $\phi^X_{0,\varepsilon}$, we get

$$\frac{\partial \phi^{-1}_{0,\varepsilon}}{\partial \varepsilon} = -\left[\phi_{0,\varepsilon}{}^* X_\varepsilon\right] \circ \phi^{-1}_{0,\varepsilon}$$

Using this equality in $\frac{d}{d\varepsilon}[f \circ \phi^{-1}_{0,\varepsilon}(z)] = [df(\phi^{-1}_{0,\varepsilon} z)] \cdot \frac{\partial \phi^{-1}_{0,\varepsilon}}{\partial \varepsilon}(z)$ we have $\frac{d}{d\varepsilon}(f \circ \phi^{-1}_{0,\varepsilon}) = -[(df) \cdot (\phi^*_{0,\varepsilon} X_\varepsilon)] \circ \phi^{-1}_{0,\varepsilon}$ or, by the very definition of Lie derivative,

$$\frac{d}{d\varepsilon}(f \circ \phi^{-1}_{0,\varepsilon}) = -\left[L_{\phi^*_{0,\varepsilon} X_\varepsilon} f\right] \circ \phi^{-1}_{0,\varepsilon}$$

Because of the equality $\varphi^*[L_X g] = L_{\varphi^* X}[\varphi^* g]$ which holds for any diffeomorphism φ and any function g, the last equation coincides with (11). ■

Remark As for the direct algorithm one can introduce operators N_s such that $\mathcal{F}_{(s)} = N_s f$, namely $N_0 = 1$, $N_s = -\sum_{j=1}^{s} \binom{s-1}{j-1} L_j N_{s-j}$. The N_s have a structure similar to the M_s. Precisely (see (22) and (23) below), for each s, M_s and N_s are sums of terms made of the same L_j but in the reversed order and with a different sign if the number of L_j is odd. For instance

$$N_1 = -L_1$$
$$N_2 = L_1^2 - L_2$$
$$N_3 = -L_1^3 + L_1L_2 + 2L_2L_1 - L_3$$
$$N_4 = L_1^4 - L_1^2L_2 - 3L_2L_1^2 - 2L_1L_2L_1 + 3L_2L_1 + 3L_2^2 + L_1L_3 - L_4$$

Relation with Giorgilli-Galgani's algorithm In [22] an apparently different algorithm to represent canonical transformations was independently introduced (not in the realm of Lie series) by Giorgilli and Galgani. It can be described in the following way. Given an ε-dependent vector field

$$\tilde{X}_\varepsilon = \sum_{s=1}^{\infty} \frac{\varepsilon^s}{s!} \tilde{X}_s \tag{17}$$

a transformation of functions $f \mapsto T_{\tilde{X}_\varepsilon} f$ is defined by

$$T_{\tilde{X}_\varepsilon} f = \sum_{s=0}^{\infty} \varepsilon^s \tilde{f}_{(s)} \tag{18}$$

with the $\tilde{f}_{(s)}$ recursively defined by (with $L_{X_j} = L_j$)

$$\begin{aligned} \tilde{f}_{(0)} &= f \\ \tilde{f}_{(s)} &= \sum_{j=1}^{s} \frac{j}{s} L_j \tilde{f}_{(s-j)} \end{aligned} \tag{19}$$

The extension to ε-dependent functions can be performed by acting linearly on Taylor series.

In fact, Giorgilli-Galgani's and Deprit's inverse transformations are two different ways of writing the same thing since, for any $\tilde{X}$ and any f, one has

$$(20) \qquad T_{\tilde{X}_\varepsilon} f = \left[\left(\phi_{0,\varepsilon}^{-\frac{\partial \tilde{X}_\varepsilon}{\partial \varepsilon}} \right)^{-1} \right]^* f$$

Namely, $T_{\tilde{X}_\varepsilon}$ coincides with the inverse of the Deprit's series generated by $X_\varepsilon = -\dfrac{\partial \tilde{X}_\varepsilon}{\partial \varepsilon}$, i.e. if for all s=1,2,... one has

$$(21) \qquad X_s = -s!\, \tilde{X}_s$$

with X_s and $\tilde{X}_s$ defined by (8) and (18). This is immediately verified since from (21), (15) and (19) it follows that (6) and (18) coincide (with $\mathcal{F}_{(s)} = s!\, \tilde{f}_{(s)}$). In the canonical case, the relation $\mathcal{X}_s = -s!\, \tilde{\mathcal{X}}_s$ also holds between the Hamilton functions of each X_s and $\tilde{X}_s$.

Closed formulas As a final topic on Deprit's Lie series we give closed formulas for the $\mathcal{F}^{(s)}$ and $\mathcal{F}_{(s)}$. As noted in [13] they can be very simply derived from the recursive formulas.

Deprit's Lie series. The expression (14) of M_s, for any $s \geqslant 1$, can be written, isolating the term with M_0,

$$M_s = L_s + \sum_{j=1}^{s-1} \binom{s-1}{j-1} M_{s-j} L_j$$

Iterating the same procedure s-1 times, so to eliminate all the M_s but $M_0 = \mathbb{1}$, one arrives at

$$M_s = L_s + \sum_{j_1=1}^{s-1} \binom{s-1}{j_1-1} L_{s-j_1} L_{j_1} +$$

$$+ \sum_{j_1=1}^{s-1} \sum_{j_2=1}^{s-1-j_1} \binom{s-1}{j_1-1}\binom{s-j_1-1}{j_2-1} L_{s-j_1-j_2} L_{j_2} L_{j_1} + \cdots\cdots +$$

$$+ \sum_{j_1=1}^{s-1} \cdots \sum_{j_{s-2}=1}^{s-1-j_1-\cdots-j_{s-1}} \binom{s-1}{j_1-1} \cdots \binom{s-1-j_1-\cdots-j_{s-1}}{j_{s-1}-1} L_{s-j_1-\cdots-j_{s-1}} L_{j_{s-1}} \cdots L_{j_1}$$

For $r=2,\ldots,s$ each sum can be rewritten, by introducing the additional summation index $j_r = s-j_1-\ldots-j_{r-1}$, as a sum on all the positive integers $j_1,\ldots,j_r$ whose sum equal s; the same is true for the term L_s. Thus one has

$$(22) \qquad M_s = \sum_{r=1}^{s} \sum_{\substack{j_1,\ldots,j_r>0 \\ j_1+\cdots+j_r=s}} \binom{s-1}{j_1-1}\binom{s-j_1-1}{j_2-1}\cdots\binom{s-j_1-\cdots-j_r-1}{j_r-1} L_{j_r}\cdots L_{j_1}$$

Inverse Deprit's series. Let us write $\mathcal{F}_{(s)} = N_s f$, with $N_0 = 1$ and $N_s = -\sum_{j=1}^{s} \binom{s-1}{j-1} L_j N_{s-j}$. Proceeding as before one finds, for all $s \geq 1$:

$$(23) \qquad N_s = \sum_{r=1}^{s} (-1)^r \sum_{\substack{j_1,\ldots,j_r>0 \\ j_1+\cdots+j_r=s}} \binom{s-1}{j_1-1}\cdots\binom{s-j_1-\cdots-j_r-1}{j_r-1} L_{j_1}\cdots L_{j_r}$$

Giorgilli-Galgani's series. Closed formulas for this case can be obtained from the last one by using the relations (18)-(19). Let X be the generator of T_X . Thus $\tilde{f}_{(s)} = s!\,\mathcal{F}_{(s)} = s!\,N_s f$, where the N_s are relative to the vector field $-\frac{\partial X}{\partial \varepsilon}$; to obtain them one has to change each L_j in $-j!\,L_j$ in (23). In this way one finds

$$\tilde{f}_{(s)} = \left\{ \sum_{r=1}^{s} \sum_{\substack{j_1,\dots,j_r>0 \\ j_1+\dots+j_r=s}} \frac{j_r j_{r-1} \cdots j_1}{j_r (j_r+j_{r-1}) \cdots (j_r+j_{r-1}+\dots+j_1)} L_{j_1} \cdots L_{j_r} \right\} f \tag{24}$$

1.3.C. Products of simple Lie series

The Taylor expansion around $\varepsilon=0$ of a composition of simple Lie transforms is easily obtained by referring to its definition (1.2.4). Expanding the function $(t_1,\dots,t_r) \longmapsto f \circ \phi_{t_1}^{X_1} \circ \dots \circ \phi_{t_r}^{X_r}$ in its (multiple) Taylor series around $t_1=\dots=t_r=0$, using the identity $\frac{\partial^{\vartheta_1+\dots+\vartheta_r}}{\partial t_1^{\vartheta_1} \dots \partial t_r^{\vartheta_r}} \left(f \circ \phi_{t_1}^{X_1} \circ \dots \circ \phi_{t_r}^{X_r} \right)_{t_1=\dots=t_r=0} = L_{X_r}^{\vartheta_r} \dots L_{X_1}^{\vartheta_1} f$, and then taking $t_1=\varepsilon,\dots,t_r=\varepsilon^r$ and collecting toghether the same powers of ε one gets

$$\phi_\varepsilon^{(X_1,\dots,X_r)} * f = \sum_{s=0}^{\infty} \frac{\varepsilon^s}{s!} \sum_{\substack{\vartheta_1,\dots,\vartheta_r \geq 0 \\ \vartheta_1+2\vartheta_2+\dots+r\vartheta_r=s}} \frac{L_{X_r}^{\vartheta_r} \cdots L_{X_1}^{\vartheta_1}}{\vartheta_r! \dots \vartheta_1!} f \tag{25}$$

This result can also be written, in a more formal way,

$$\phi_\varepsilon^{(X_1,\dots,X_r)} * f = (\phi_{\varepsilon^r}^{X_r})^* \cdots (\phi_\varepsilon^{X_1})^* f = e^{\varepsilon^r L_{X_r}} \cdots e^{\varepsilon L_{X_1}} f$$

which justifies the name of product of Lie series usually employed for this algorithm.

Let us introduce some notation to be used hereafter. For all the integers $s \geq 0$ and $r > 0$ we put

$$P(s;r) = \left\{ \vartheta = (\vartheta_1,\dots,\vartheta_r) \in \mathbb{N}_0^r : \sum_{j=1}^{r} j\vartheta_j = s \right\} \tag{26}$$

where $\mathbb{N}_0$ is the set of all the nonnegative integers. With reference to a collection $X_1,\dots,X_r$ of vector fields we put $L_j = L_{X_j}$ and write (25) in the form

$$\phi_\varepsilon^{(X_1,\dots,X_r)} * f = \sum_{s=0}^{\infty} \varepsilon^s f_{(s)} \tag{27}$$

where, for each $s \geq 0$,

$$f_{(s)} = \sum_{\nu \in P(s;r)} \frac{L_r^{\nu_r} \cdots L_1^{\nu_1}}{\nu_r! \cdots \nu_1!} f \tag{28}$$

Remarks (i) Equations (26) and (28) hold also when some of the vector fields $X_1,\dots,X_r$ is null, with the convention that $L^{\nu}_{X=0}$ $=0$ for all $\nu>0$ but $L^{0}_{X=0} = 1$.

(ii) There is no ambiguity in writing, for $s \geq 1$,

$$f_{(s)} = \sum_{\nu \in P(s)} \frac{L_s^{\nu_s} \cdots L_1^{\nu_1}}{\nu_s! \cdots \nu_1!} f \tag{28'}$$

with $P(s)=P(s;s)$. Indeed, for $1 \leq s \leq r$ it is $P(s;r)=P(s;s)$ while when $s>r$ one can simply define $X_{r+1} = \dots = X_s = 0$. When $s=0$, equation (28') holds if the convention $P(0) = P(0;r)$, for any $r \geq 1$, is made. We will often use this form instead than (28).

(iii) The transformation of ε-dependent functions can be obviously performed by acting linearly on Taylor series. In the canonical case the L_j can also be interpreted as Poisson brackets.

The closed formula (28) shows that each $f_{(s)}$ is obtained by acting on f only with "ordered" compositions of the operators

$L_1,\dots,L_r$, in the sense that only those compositions $L_{i_p} L_{i_{p-1}} \dots L_{i_1}$ with $i_p \geq i_{p-1} \geq \dots \geq i_1$ are present. This is peculiar of products of Lie series and is the main difference with the various variants of Deprit's series (see (22)-(24)).

It is also possible to give a recursive formula for the $f_{(s)}$; it turns out to be exactly like that of Giorgilli-Galgani's algorithm, except for the order in which the operators L_j appear.

To derive it let us introduce some terminology. With reference to a collection $X_1,\dots,X_r$ of vector fields we introduce an "ordering operator" $:\;:$ on the space of all the differential monomials $L_{i_p}\dots L_{i_1}$ ($p=1,2,\dots$; $i_1,\dots,i_p=1,\dots,r$) defined by $:L_{i_p}\dots L_{i_1}: \;= L_{\sigma(i_p)}\dots L_{\sigma(i_1)}$ where σ is any permutation of $\{i_1,\dots,i_p\}$ such that $\sigma(i_p) \geq \sigma(i_{p-1}) \geq \dots \geq \sigma(i_1)$; for instance $:L_1 L_3 L_2: \;= L_3 L_2 L_1$. A monomial is then said ordered if it is invariant under $:\;:$. By linearity, the ordering is extended to linear combinations of monomials. Finally, for any function f, we put $:L_{i_p}\dots L_{i_1} f: \;\equiv \;:L_{i_p}\dots L_{i_1}: f$. Note that all this also hold when $r = \infty$.

Recursive formula. The functions $f_{(s)}$ ($s=0,1,\dots$) defined as in (28) satisfy

$$f_{(0)} = f \tag{29}$$

$$f_{(s)} = \sum_{j=1}^{s} \frac{j}{s} :L_j f_{(s-j)}:$$

Proof Consider any $s\geq 1$ and assume $r\geq s$ (otherwise define $X_{r+1} = \ldots = X_s = 0$). Using (28') and the definition of the ordering we have, for each $j=1,\ldots,s$:

$$:L_j f_{(s-j)}: \;=\; \sum_{\vartheta\in P(s-j)} \frac{L_{s-j}^{\vartheta_{s-j}} \cdots L_j^{1+\vartheta_j} \cdots L_1^{\vartheta_1}}{\vartheta_{s-j}!\cdots \vartheta_j! \cdots \vartheta_1!}\, f$$

We now perform the change of summation indexes $\vartheta = (\vartheta_1,\ldots,\vartheta_r) \mapsto \vartheta' = (\vartheta'_1,\ldots,\vartheta'_r)$ with $\vartheta'_j = 1+\vartheta_j$, $\vartheta'_i = \vartheta_i$ for $i<j$ and $\vartheta'_i = 0$ for $i>j$. Since then $\vartheta\in P(s-j)$ iff $\vartheta'\in P(s)$ and $\vartheta'_j\geq 1$, we have

$$:L_j f_{(s-j)}: \;=\; \sum_{\substack{\vartheta'\in P(s)\\ \vartheta'_j\geq 1}} \vartheta'_j\, \frac{L_s^{\vartheta'_s}\cdots L_j^{\vartheta'_j}\cdots L_1^{\vartheta'_1}}{\vartheta'_s!\cdots\vartheta'_j!\cdots\vartheta'_1!}$$

But here the limitation $\vartheta'_j\geq 1$ can be dropped out because of the factor ϑ'_j in each addend. Thus, summing on j, we obtain

$$\sum_{j=1}^{s} \frac{j}{s} :L_j f_{(s-j)}: \;=\; \sum_{j=1}^{s} \sum_{\vartheta\in P(s)} \frac{j\vartheta_j}{s}\, \frac{L_s^{\vartheta_s}\cdots L_1^{\vartheta_1}}{\vartheta_s!\cdots\vartheta_1!}\, f$$

Exchanging the two sums at the r.h.s. and using $\sum_{j=1}^{s} j\vartheta_j = s$ this reduces to (28'). ∎

Remarks (i) As for Deprit's Direct series, this recursive formula is more suitable to symbolically generate the differential operators defining the $f_{(s)}$ than to recursively compute them.

(ii) There is obviously a strict formal analogy between the recursive formulas for the products of Lie series and Giorgilli-Galgani's algorithm. Indeed, consider a sequence X_1, X_2,.. of vector fields and construct $\phi_\varepsilon^{(X_1,X_2,\ldots)*} f = \sum_{s\geq 0} \varepsilon^s f_{(s)}$;

with the same $X_1, X_2, ..$ construct $X_\varepsilon = \sum_{s\geq 1} \varepsilon^s X_s$ and then $T_{X_\varepsilon} f = \sum_{s\geq 1} \varepsilon^s \tilde{f}_{(s)}$. It then immediately follows from (19) and (29) that, for all $s\geq 0$, one has $f_{(s)} = :\tilde{f}_{(s)}:$. Namely, if in the expression (24) of $\tilde{f}_{(s)}$ all the monomials $L_{j_r}...L_{j_1}$ are substituted with their ordered monomials, then the $f_{(s)}$ defined by (28) results (similar analogies also exist with Deprit's direct and inverse series -but not so sharp, because different numerical coefficients arise). Clearly, this is only a formal analogy between the differential operators defining $f_{(s)}$ and $\tilde{f}_{(s)}$, since they are in general completly different functions, and so are the two transformations.

1.4 EXISTENCE OF LIE TRANSFORMS

We give in this section some results on the analyticity of Lie transforms, which are in turn convergence results for the corresponding Lie series. For shortness we treat only the canonical case, in a form suitable for the applications to Hamiltonian perturbation theory; the noncanonical case could be treated along the same lines (actually, with some minor simplifications). The analyticity of (complex) flows will be simply inferred from that of the vector fields (or the Hamilton functions) by means of general theorems on differential equations; this procedure has been used, for simple Lie series, in [7,8,9].

Let us note that there exists a different, but equivalent, approach in which one directly proves the convergence of the Lie series and infer from this the analyticity of the flow; see [24,22,35].

1.4.A. Preliminaries

We consider only the complex analytic case. This is common in (part of) perturbation theory because the derivatives of complex analytic functions can be controlled with the functions themselves by means of Cauchy inequalities. The parameter ε is assumed complex, too. We first introduce some notation with reference to a space C^m and then specialize to the canonical case.

Norms and notations. In C^m we use the maximum coordinate norm, i.e. $\|z\| = \max|z_i|$ if $z=(z_1,\ldots,z_m)\in C^m$. For any $\alpha>0$, the polydisk $\{z'\in C^m : \|z'-z\|\leq\alpha\}$ will be called a α-neighbourhood of z. If $D\subset C^m$ and $\alpha>0$, $D+\alpha$ denotes the union of all the α-neighbourhoods of the points of D, while $D-\alpha$ is the set of those points which are in D toghether with a whole α-neighbourhood (it will be always understood that α is so small that $D-\alpha$ is not empty). The symbol B_T always denotes the disk $\{t\in C : |t|\leq T\}$.

The norm used for functions is that of the supremum. If $D\subset C^m$ we write $\|f\|_D=\sup_{z\in D}|f(z)|$. For the restriction of a vector field $X=(X_1,\ldots,X_m)$ to some $D\subset C^m$ the norm is $\|X\|_D = \max_i \sup_{z\in D}|X^i(z)|$.

Similarly, the maximum of the supremum of the moduli of the components is used as norm for all the tensor fields (like multiple derivatives). When $X:(t,z)\mapsto X_t(z)$ is a nonconservative vector field defined on some set $B_T\times D\subset C^{m+1}$ we write $\|X\|_{B_T\times D} = \sup_{t\in B_T}\|X_t\|_D$.

Cauchy inequalities Let $f:B_T\to C$ be an analytic function. Then, for all $s\geq 0$ it is

$$(1)\qquad \left|\frac{d^s f}{dt^s}(0)\right| \leq \frac{s!}{T^s}\sup_{|t|<T}|f(t)| = \frac{s!}{T^s}\|f\|_{B_T}$$

This well known inequality, which immediately follows from Cauchy's integral formula, is the prototype of the so called Cauchy inequalities and is easily generalized as follows. For a multi-index $\nu=(\nu_1,\dots,\nu_m)$, i.e. a vector with nonnegative integers components, we write $|\nu|\equiv\sum_{j=1}^{m}\nu_j$, $\nu!=\prod_{j=1}^{m}(\nu_j!)$ and $\frac{\partial^{|\nu|}}{\partial z^\nu}=\frac{\partial^{\nu_1+\dots+\nu_m}}{\partial z_1^{\nu_1}\dots\partial z_m^{\nu_m}}$. Thus, for any function f analytic on a domain $D\subset C^m$, all multiindexes ν and all $s\geq 0$, one has

$$(2)\qquad \left\|\frac{\partial^{|\nu|}f}{\partial z^\nu}\right\|_{D-\alpha} \leq \frac{\nu!}{\alpha^{|\nu|}}\|f\|_D$$

The Hamiltonian case. When dealing with nearly integrable Hamiltonian systems we will be primarily concerned with real analytic functions $f:B\times T^n\to R$, $(p,q)\mapsto f(p,q)$, $B\subset R^n$ being some open set. To apply Cauchy inequalities we will need analytic extensions of these functions to some complex neighbourhood of $B\times T^n$. Let us treat the angles $q\in T^n$ as real variables $q\in R^n$, by (tacitly) imposing to the functions of being 2π-periodic in them, and consider complex neighbourhoods of $B\times T^n$ of the following type: for any $\rho>0$, $\sigma>0$ we define

(3) $$D_{\rho,\sigma}(B) = \{(p,q)\in C^{2n} : d(p,B)\leq\rho \ , \|Im(q)\|<\sigma\}$$

Here $d(p,B) = \inf\limits_{p'\in B}\|p-p'\|$ is the distance of $p\in C^n$ from B, regarded as a subset of C^n and $\|\cdot\|$ is the maximum coordinate norm. Note that

(3') $$D_{\rho,\sigma}(B) = D_\rho(B)\times\mathcal{S}_\sigma$$

with $D_\rho(B)=\{p\in C^n: \ d(p,B)\leq\rho\}$ and $\mathcal{S}_\sigma = \{q\in C^n : \|Im(q)\|<\sigma\}$.
When no confusion arises the sets $D_{\rho,\sigma}(B)$ will be often denoted $D_{\rho,\sigma}$ or simply D. Furthermore, for any $0<\delta<\rho$, $0<\xi<\sigma$ we define

(4) $$D_{\rho,\sigma}(B)-(\delta,\xi) = D_{\rho-\delta,\sigma-\xi}(B-\delta)$$

and, for all $\delta>0$, $\xi>0$:

(5) $$D_{\rho,\sigma}(B)+(\delta,\xi) = D_{\rho+\delta,\sigma+\xi}(B+\delta)$$

where $B\pm\delta$ are defined as before (these changes in B are completely irrelevant for the present treatment but will be useful in the next chapter). If the c_i are positive numbers, we usually write $D\pm\sum\limits_i c_i(\delta_i,\xi_i)$ to mean $D\pm(\sum\limits_i c_i\delta_i \ , \sum\limits_i c_i\xi_i)$.
For an analytic function $f:D\rightarrow C$ (with $D=D_{\rho,\sigma}(B)$), Cauchy

inequalities are, for all $0<\delta<\rho$, $0<\xi<\sigma$ and all multiindexes μ and ν:

$$\left\| \frac{\partial^{|\nu|} f}{\partial p^{\nu}} \right\|_{D-(\delta,0)} \leq \frac{\nu!}{\delta^{|\nu|}} \|f\|_D$$

$$\left\| \frac{\partial^{|\mu|} f}{\partial q^{\mu}} \right\|_{D-(0,\xi)} \leq \frac{\mu!}{\xi^{|\mu|}} \|f\|_D \tag{6}$$

$$\left\| \frac{\partial^{|\mu|+|\nu|} f}{\partial p^{\nu} \partial q^{\mu}} \right\|_{D-(\delta,\xi)} \leq \frac{\mu!\,\nu!}{\delta^{|\mu|}\,\xi^{|\nu|}} \|f\|_D$$

Analitycity of complex flows. We give now some quantitative estimates on the existence of complex analytic (nonautonomous) flows which constitute the basis of the following work. For shortness we consider only a canonical flow defined on a set $D_{\rho,\sigma}(B)$ and give the estimates in terms of a Hamilton function of the Hamiltonian vector field which induces the flow. The noncanonical case could be treated analogously, with some simplifications (see point (a) of the following proof).

Let us introduce some terminology. We will say that a mapping $\phi : D \subset \mathbb{C}^m \to \mathbb{C}^m$ is α-close to the identity iff $\|\phi(z)-z\| \leq \alpha$ for all $z \in D$. This obviously implies $\phi(D) \subset D+\alpha$; moreover, when ϕ is a diffeomorphism of D onto its image, this also implies that $D-\alpha \subset \phi(D)$. When D is a set $D_{\rho,\sigma}(B)$, by saying that ϕ is $\alpha(\delta,\xi)$-close to the identity we mean that, for all $(p,q) \in D$, one has, with $(p',q') = \phi(p,q)$, $\|p-p'\| \leq \alpha\delta$, $\|\mathrm{Im}(q-q')\| \leq \alpha\xi$; then $\phi(D) \subset D+\alpha(\delta,\xi)$ and, if ϕ is a diffeomorphism, $D-\alpha(\delta,\xi) \subset \phi(D)$.

Lemma1.1 Let $B \subset R^n$ be an open set and ρ, σ, a positive numbers; write $D = D_{\rho,\sigma}(B)$. Consider an analytic function $\mathcal{X}:(p,q,t) \mapsto \mathcal{X}_t(p,q)$ defined on $B_a \times D$. Consider positive numbers $\delta < \rho$, $\xi < \sigma$ and define $T^* = \min(a/8,\ \delta\xi/8\|\mathcal{X}\|_{B_a \times D})$. Then
(i) $\phi^{\mathcal{X}}_{0,\cdot} : B_{2T^*} \times D-(\delta,\xi) \longrightarrow D-(\delta/2,\xi/2)$ exists and is analytic;
(ii) for each $|t|<T^*$, $\phi^{\mathcal{X}}_{0,t}$ is a canonical analytic diffeomorphism of $D-(\delta,\xi)$ onto its image and it is $(t/4T^*)(\delta,\xi)$-close to the identity;
(iii) if $\mathcal{X}$ does not depend on t, the same is true with T replaced by $T^* = \delta\xi/8\|\mathcal{X}\|_D$.

Proof (a) Let us first recall some general results on differential equations. Let $Y=(Y_1,\dots,Y_m)$ be a vector field analytic on a neighbourhood of an open set $W \subset C^m$ $(m \geq 1)$. Consider m positive numbers $(\beta_1,\dots,\beta_m)=\vec{\beta}$ and the set

$$W-\vec{\beta} = \{ z \in W: \text{ if } |z'_i - z_i| \leq \beta_i \text{ for } i=1,\dots,m \text{ then } z' \in W \}.$$

Note that, for all $z \in W-\vec{\beta}$, there is the apriori estimate

$$\phi^Y_t z \in W \qquad \text{for all } |t| < S = \min(\beta_i / \|Y_i\|_W).$$

It then immediately follows from the standard existence-uniqueness theorem (proved with Picard's successive approximations method, as for instance in [28]) the existence, uniqueness and analyticity of the integral curve $t \mapsto \phi^Y_t z$ of Y through z for all $z \in W-\vec{\beta}$ and $|t|<S$. Toghether with the theorem on the differential dependence on the initial conditions (as can be

found for instance in [30]), this implies that $\phi^{Y}: B_S \times W-\vec{\beta} \longrightarrow W$ is analytic. The (analytic) invertibility of ϕ_t^{Y} from from $W-\vec{\beta}$ onto its image for all $|t|<S/2$ is easily proved in a standard way (see ([1]). Finally, for all $z \in W-\vec{\beta}$ and $|t|<S$ it is

$$\|(\phi^{Y} z)_i - z_i\| \leq |t| \cdot \|Y_i\|_W = \frac{|t|}{S}\beta_i$$

so that ϕ_t^{Y} is $\frac{|t|}{S}\min(\beta_i)$-close to the identity.

(b) We now apply these results to the case of this lemma, firstly assuming $\frac{\partial \mathcal{X}}{\partial t}=0$, as in (iii). Note that the Hamiltonian vector field $X = (-\frac{\partial \mathcal{X}}{\partial q}, \frac{\partial \mathcal{X}}{\partial p})$ is analytic on D and then bounded on any subset of D (the unboundedness of D in the "angle" directions does not evidently play here any role; the restriction of B in $D-(\delta, \xi)$ is completely irrelevant, too). Take $\delta_1, \delta_2, \xi_1, \xi_2 > 0$ such that $\delta_1+\delta_2 = \delta < \rho$, $\xi_1+\xi_2 = \xi < \sigma$. Proceed now as before, by noting that, for all $(p,q) \in D-(\delta,\xi)$, there is the apriori estimate $\phi_t^{X}(p,q) \in D-(\delta_1, \xi_1)$ for all

$$|t| < S = \min(\delta_2 / \|\tfrac{\partial \mathcal{X}}{\partial q}\|_{D-(\delta_1,\xi_1)}, \xi_2/\|\tfrac{\partial \mathcal{X}}{\partial p}\|_{D-(\delta_1,\xi_1)}).$$

It then follows that $\phi^{X}: B_S \times D-(\delta,\xi) \longrightarrow D-(\delta_1, \xi_1)$ is analytic and that, for all $|t|<S/2$, ϕ_t^{X} is a $(|t|/S)(\delta_2, \xi_2)$-close to the identity analytic diffeomorphism of $D-(\delta,\xi)$ onto its image. Cauchy inequalities (6) give

$$\|\frac{\partial \mathcal{X}}{\partial q}\|_{D-(\delta_1,\xi_1)} \leq \frac{1}{\xi_1}\|\mathcal{X}\|_D \quad , \quad \|\frac{\partial \mathcal{X}}{\partial p}\|_{D-(\delta_1,\xi_1)} \leq \frac{1}{\delta_1}\|\mathcal{X}\|_D$$

so that $S \geq \min(\delta_1\xi_2, \delta_2\xi_1)/\|\mathcal{X}\|_D = 2T$ and then these results also hold with S replaced by 2T. To optimize this lower bound take $\delta_1 = \delta_2 = \delta/2$, $\xi_1 = \xi_2 = \xi/2$ so that $T = \delta\xi/8\|\chi\|_D = T^*$ is obtained. Finally, the canonicity of ϕ_t^{χ} is obvious.

(c) To treat the case in which χ depends also on t, pass to the extended space C^{2n+1} with coordinates (t,p,q). Proceeding as in (a), take $W = B_{a/2} \times D-(\delta/2, \xi/2)$, $Y=(1,-\frac{\partial\chi}{\partial q}, \frac{\partial\chi}{\partial p})$, $\beta_1 = a/4$, $\beta_j = \delta/2$ for $1 \leq j \leq n+1$, $\beta_j = \xi/2$ for $n+2 \leq j \leq 2n+1$ (the restriction of B_a to $B_{a/2}$ in defining W woold be not strictly necessary, but it assures the boundedness of Y on W). It follows that

$$S = \min\left(a/4, \delta/2\left\|\frac{\partial\chi}{\partial q}\right\|_{B_{a/2}\times D-(\delta/2,\xi/2)}, \xi/2\left\|\frac{\partial\chi}{\partial p}\right\|_{B_{a/2}\times D-(\delta/2,\xi/2)}\right)$$

$$\geq \min\left(a/4, \delta\xi/4\|\chi\|_{B_{a/2}\times D}\right) = 2T^*.$$

Thus, applying the results of (a) to the flow $\phi^Y : B_{2T^*}\times B_{a/4}\times D-(\delta,\xi) \longrightarrow B_{a/2}\times D-(\delta/2, \xi/2)$ and using the fact that $\phi^Y(s,(t,p,q)) = (t+s, \phi^{\chi}_{t,t+s}(p,q))$ one easily proves all the assertions. ∎

1.4.B Estimates on Lie transforms and Lie series

We give now results on the analyticity of Lie transforms, and thus on the convergence of Lie series. For shortness, we consider only the canonical case and sets of the type $D_{\rho,\sigma}(B)$; the estimates are given with reference to the generating functions rather than to the generating vector fields, as this is useful in Hamiltonian perturbation theory. Besides lower bounds for the range of existence we also give bounds on the errors accomplished when truncating the series.

Proposition 1.1 Let $B \subset R^n$ be an open set, ρ and σ positive numbers and $\mathcal{X}$ a function analytic on $D = D_{\rho,\sigma}(B)$. Consider positive numbers $\delta < \rho$, $\xi < \sigma$ and define

$$\varepsilon_1^* = \frac{\delta \xi}{2^3 \|\mathcal{X}\|_D} \tag{7}$$

Then for each $|\varepsilon| < \varepsilon_1^*$ and any function f analytic on D:

(i) $\phi_\varepsilon^{\mathcal{X}}$ is a $\frac{|\varepsilon|}{4\varepsilon_1^*}(\delta,\xi)$-close to the identity analytic canonical diffeomorphism of $D-(\delta,\xi)$ onto its image;

(ii) $f \circ \phi_\varepsilon^{\mathcal{X}} = \exp(\varepsilon L_{\mathcal{X}}) f$ is analytic on $D-(\delta,\xi)$ and

$$\| f \circ \phi_\varepsilon^{\mathcal{X}} \|_{D-(\delta,\xi)} \leq \|f\|_D$$

(iii) for any $p \geq 0$, the p-th remainder of the Lie series, $R(p,\varepsilon)f = \sum_{s=p}^{\infty} \frac{\varepsilon^s}{s!} L_{\mathcal{X}}^s f$ is bounded by

$$\| R(p,\varepsilon) f \|_{D-(\delta,\xi)} \leq \|f\|_D \frac{\left(\frac{|\varepsilon|}{2\varepsilon_1^*}\right)^p}{1 - \frac{|\varepsilon|}{2\varepsilon_1^*}} \leq 2 \|f\|_D \left(\frac{|\varepsilon|}{2\varepsilon_1^*}\right)^p \tag{8}$$

Proof By Lemma 1.1 $\phi^{\mathcal{X}}: B_{2\varepsilon_1^*} \times D-(\delta,\xi) \to D$ is analytic and satisfies (i). In turn, this implies (ii). Applying Cauchy inequality (1) to the function $\varepsilon \mapsto f \circ \phi_\varepsilon^{\mathcal{X}}(p,q)$ we get, for each $(p,q) \in D-(\delta,\xi)$:

$$\left| \frac{d^s}{dt^s} f \circ \phi_\varepsilon^{\mathcal{X}}(p,q) \right|_{\varepsilon=0} \leq \frac{s!}{(2\varepsilon_1^*)^s} \sup_{|\varepsilon|<2\varepsilon_1^*} \left| f \circ \phi_{0,\varepsilon}^{\mathcal{X}}(p,q) \right| \leq \frac{s!}{(2\varepsilon_1^*)^s} \|f\|_{D-(\frac{\delta}{2},\frac{\xi}{2})}$$

So $\| L^s_{\mathcal{X}} f \|_{D-(\delta,\xi)} \leq \frac{s!}{(2\varepsilon^*_1)^s} \| f \|_D$ and then

$$\sum_{s=p}^{\infty} \| \frac{\varepsilon^s}{s!} L^s_{\mathcal{X}} f \|_{D-(\delta,\xi)} \leq \| f \|_D \sum_{s=p}^{\infty} \left(\frac{|\varepsilon|}{2\varepsilon^*_1} \right)^s$$

from which (8) follows by elementary properties of geometric series; the last inequality in (8) is obtained by noting that $(1-x)^{-1} \leq 2$ for all $0 < x \leq 1/2$. ∎

We consider now the case of a composition of Lie transforms.

Proposition 1.2 Let $B \subset R^n$ be an open set, ρ and σ positive numbers and denote $D = D_{\rho,\sigma}(B)$. Assume that the functions $\mathcal{X}_1, \dots, \mathcal{X}_r$ (any $r \in \mathbb{N}$) are analytic on D and satisfy $\| \mathcal{X}_s \|_D \leq A^s$ for some $A \geq 0$ and all $1 \leq s \leq r$. Consider any δ and ξ such that $0 < \delta < \min(\rho, 1)$, $0 < \xi < \min(\sigma, 1)$ and define

$$\text{(9)} \qquad \varepsilon^*_r = \frac{\delta \xi}{2^5 A}$$

Then, for each $|\varepsilon| < \varepsilon^*_r$ and any function f analytic on D:

(i) $\phi^{\mathcal{X}_1}_{\varepsilon} \circ \dots \circ \phi^{\mathcal{X}_r}_{\varepsilon^r}$ is an analytic canonical diffeomorphism $\frac{|\varepsilon|}{4\varepsilon^*_r}$ (δ, ξ)-close to the identity of $D-(\delta,\xi)$ onto its image;

(ii) $f \circ \phi^{\mathcal{X}_1}_{\varepsilon} \circ \dots \circ \phi^{\mathcal{X}_r}_{\varepsilon^r} = \exp(\varepsilon^r L_{\mathcal{X}_r}) \cdots \exp(\varepsilon L_{\mathcal{X}_1}) f = \sum_{s \geq 0} \varepsilon^s f_{(s)}$ (see (1.3.28)) is analytic on $D-(\delta,\xi)$ and

$$\| f \circ \phi^{\mathcal{X}_1}_{\varepsilon} \circ \dots \circ \phi^{\mathcal{X}_r}_{\varepsilon^r} \|_{D-(\delta,\xi)} \leq \| f \|_D ;$$

(iii) for any $p \geq 0$, the p-th remainder $R(p,\varepsilon) f = \sum_{s=p}^{\infty} \varepsilon^s f_{(s)}$ is bounded by

$$\text{(10)} \qquad \| R(\varepsilon,p) f \|_{D-(\delta,\xi)} \leq \| f \|_D \frac{\left(\frac{|\varepsilon|}{2\varepsilon^*_r} \right)^p}{1 - \frac{|\varepsilon|}{2\varepsilon^*_r}} \leq 2 \| f \|_D \left(\frac{|\varepsilon|}{2\varepsilon^*_r} \right)^p$$

Proof Consider any $s\geq 1$ and any set $D' \subset D-(\delta/2^s,\xi/2^s)$. Define $T_s = \delta\xi/2^{3+2s}\|\mathcal{X}_s\|_D$. It follows from lemma 1.1 (with δ,ξ replaced by $\delta/2^s$, $\xi/2^s$) that $\phi^{\mathcal{X}_s}: B_{2T_s}\times D' \to D$ is analytic and, for each $|t_s|<T_s$, $\phi^{\mathcal{X}_s}_{t_s}$ is a $(t_s/4T_s)(\delta/2^s,\xi/2^s)$-closed to the identity canonical diffeomorphism of D' onto its image. Define now $\tilde{\phi}^{\mathcal{X}_s}:(\varepsilon,p,q)\mapsto \phi^{\mathcal{X}_s}_{\varepsilon^s}(p,q)$ and note that, being $\|\mathcal{X}_s\|_D\leq A^s$ and $\delta\xi<1$, it is $2T_s \geq \delta\xi/2^{2+2s}A^s \geq (\delta\xi/2^4A)^s = (2\varepsilon^*_p)^s$ and $T_s \geq (\varepsilon^*_p)^s$. So it follows that $\tilde{\phi}^{\mathcal{X}_s}: B_{2\varepsilon^*_p}\times D'\to D$ is analytic and, for each $|\varepsilon|<\varepsilon^*_p$, $\phi^{\mathcal{X}_s}_{\varepsilon^s}$ is a $(1/4)(|\varepsilon|/2\varepsilon^*_p)^s(\delta,\xi)$-close to the identity analytic canonical diffeomorphism of D'. Since for $|\varepsilon|<\varepsilon^*_p$ it is $\sum_{s=1}^{\infty}\frac{1}{4}(|\varepsilon|/2\varepsilon^*_p)^s \leq |\varepsilon|/4\varepsilon^*_p$, we get that $\phi^{\mathcal{X}_1}\circ\dots\circ\phi^{\mathcal{X}_r}: B_{2\varepsilon^*_p}\times D-(\delta,\xi)\to D$ is analytic and, moreover, that, for each $|\varepsilon|<\varepsilon^*_p$, $\phi^{\mathcal{X}_1}_{\varepsilon}\circ\dots\circ\phi^{\mathcal{X}_r}_{\varepsilon^r}$ is a $(|\varepsilon|/4\varepsilon^*_p)(\delta,\xi)$-close to the identity analytic diffeomorphism of $D-(\delta,\xi)$ onto its image. So, (i) and (ii) follow. For (iii), using the analyticity of $\varepsilon\mapsto f\circ\phi^{\mathcal{X}_1}_{\varepsilon}\circ\dots\circ\phi^{\mathcal{X}_r}_{\varepsilon^r}$ for $|\varepsilon|<2\varepsilon^*_p$, the definition (1.3.27) of $f_{(s)}$ and Cauchy inequality, one gets that for all $(p,q)\in D-(\delta,\xi)$ and all $s\geq 0$ it is

$$|f_{(s)}(p,q)| = \frac{1}{s!}\left|\frac{d^s}{dt^s}\left[f\circ\phi^{\mathcal{X}_1}_{\varepsilon}\circ\dots\circ\phi^{\mathcal{X}_r}_{\varepsilon^r}(p,q)\right]_{\varepsilon=0}\right| \leq \frac{1}{(2\varepsilon^*_p)^s}\|f\|_D$$

from which (10) immediately follows as in the previous case. ∎

We finally consider Deprit's direct and inverse series, assuming a knowledge of the Taylor coefficients $\mathcal{X}_s = \left(\frac{d^{s-1}\mathcal{X}_\varepsilon}{d\varepsilon^{s-1}}\right)_{\varepsilon=0}$ of the generating function $\mathcal{X}_\varepsilon$ rather than of $\mathcal{X}_\varepsilon$ itself; this is irrelevant but more close to applications, where the $\mathcal{X}_s$ are constructed.

Proposition 1.3 Let $B \subset R^m$ be an open set, ρ and σ positive numbers. Consider functions $\mathcal{X}_s$ $(s=1,2,..)$ which are analytic on $D = D_{\rho,\sigma}(B)$ and satisfy $\|\mathcal{X}_s\|_D \leq (s-1)!A^s$ for some $A > 0$. Define the function $\mathcal{X}_\varepsilon = \sum_{s\geqslant 0} \frac{\varepsilon^s}{s!} \mathcal{X}_{s+1}$. Consider any δ and ξ such that $0 < \delta < \min(\rho, 1)$, $0 < \xi < \min(\sigma, 1)$ and define

$$(11) \qquad \varepsilon^*_D = \frac{\delta\,\xi}{2^4 A}$$

Then for each $|\varepsilon| < \varepsilon^*_D$ and any function f analytic on D:

(i) $\phi^{\mathcal{X}}_{0,\varepsilon}$ is an analytic canonical diffeomorphism $\frac{|\varepsilon|}{4\varepsilon^*_D}$ (δ,ξ)-close to the identity of $D-(\delta,\xi)$ onto its image;

(ii) $f\circ\phi^{\mathcal{X}}_{0,\varepsilon} = \sum_{s\geqslant 0} \frac{\varepsilon^s}{s!} f^{(s)}$ (see (1.3.6)) is analytic on $D-(\delta,\xi)$ and

$$\|f\circ\phi^{\mathcal{X}}_\varepsilon\|_{D-(\delta,\xi)} \leqslant \|f\|_D ;$$

(iii) for each p 0, the p-th remainder $R(\varepsilon,p)f = \sum_{s\geqslant p} \frac{\varepsilon^s}{s!} f^{(s)}$ is bounded by

$$(12) \qquad \|R(\varepsilon,p)f\|_{D-(\delta,\xi)} \leqslant \|f\|_D \frac{\left(\frac{|\varepsilon|}{2\varepsilon^*_D}\right)^p}{1-\frac{|\varepsilon|}{2\varepsilon^*_D}} \leqslant 2\,\|f\|_D \left(\frac{|\varepsilon|}{2\varepsilon^*_D}\right)^p$$

Proof For all $|\varepsilon| < 1/A$ it is $\sum_{s=0}^{\infty} \|\frac{\varepsilon^s}{s!}\mathcal{X}_{s+1}\|_D \leqslant A(1-|\varepsilon|A)^{-1}$, so that $\mathcal{X}$ is analytic on $B_{1/A}\times D$ and satisfies $\|\mathcal{X}\|_{D\times B_{1/2A}} \leqslant 2A$. Define $T = \min(1/8A,\ \delta\xi/8\,\|\mathcal{X}\|_{D\times B_{1/2A}})$; it then follows from lemma 1.1 that $\phi^{\mathcal{X}}_{0,\cdot}: B_{2T}\times D-(\delta,\xi) \longrightarrow D$ is analytic and, for each $|\varepsilon|<T$, $\phi^{\mathcal{X}}_{0,\varepsilon}$ is an analytic canonical diffeomorphism $(\varepsilon/4T)$ (δ,ξ)-close to the identity of $D-(\delta,\xi)$. Since $T \geqslant \min(1/8A,$

$\delta\xi/16A) = \delta\xi/16A = \varepsilon_D^*$, (i) and (ii) follow. To estimate the remainder apply Cauchy inequality to $\varepsilon \longmapsto f\circ\phi_{0,\varepsilon}^{\mathcal{X}}(p,q)$, which is analytic for $|\varepsilon| < 2\varepsilon_D^*$ and $(p,q)\in D-(\delta,\xi)$, and proceed as in the proof of proposition 1.1. ∎

Proposition 1.4 In the same hypothesis of proposition 1.3, define

$$(13) \qquad \varepsilon_I^* = \frac{\delta\xi}{2^5 A}$$

Then for each $|\varepsilon| < \varepsilon_I^*$ and any function f analytic on D:

(i) $(\phi_{0,\varepsilon}^{\mathcal{X}})^{-1}$ is a $\frac{|\varepsilon|}{4\varepsilon_I^*}(\delta,\xi)$-close to the identity analytic canonical diffeomorphism of $D-(\delta,\xi)$ onto its image;

(ii) $f\circ(\phi_{0,\varepsilon}^{\mathcal{X}})^{-1} = \sum_{s\geq 0}\frac{\varepsilon^s}{s!}\mathcal{J}_{(s)}$ (see (1.3.10)) is analytic on $D-(\delta,\xi)$ and

$$\|f\circ(\phi_{0,\varepsilon}^{\mathcal{X}})^{-1}\|_{D-(\delta,\xi)} \leq \|f\|_D \quad ;$$

(iii) for all $p \geq 0$, the p-th remainder $R(\varepsilon,p)f = \sum_{s\geq p}\frac{\varepsilon^s}{s!}\mathcal{J}_{(s)}$ is bounded by

$$(14) \qquad \|R(\varepsilon,p)f\|_{D-(\delta,\xi)} \leq \|f\|_D \frac{\left(\frac{|\varepsilon|}{\varepsilon_I^*}\right)^p}{1-\frac{|\varepsilon|}{\varepsilon_I^*}}$$

Proof According to the previous proposition, for each $|\varepsilon| < \varepsilon_D^* = \delta\xi/2^4A$, $(\phi_{0,\varepsilon}^{\mathcal{X}})^{-1}$ is an analytic diffeomorphism $(\varepsilon/4\varepsilon_D^*)(\delta,\xi)$-close to the identity of $D-(5\varepsilon/4\varepsilon_D^*)(\delta,\xi)$ onto its image, and then of $D-\frac{5}{4}(\delta,\xi)$, too. So, $\varepsilon \mapsto f\circ(\phi^{\mathcal{X}}_{0,\varepsilon})^{-1}(p,q)$ is analytic for $|\varepsilon|<\varepsilon^*_D$ and $(p,q)\in D-\frac{5}{4}(\delta,\xi)$; it follows that

$$\left\| \frac{d^s}{dt^s} f\circ(\phi^{\mathcal{X}}_{0,\varepsilon})^{-1} \right\|_{D-(\delta,\xi)} \leq \frac{s!}{(\varepsilon^*_D)^s} \|f\|_D$$

so that

$$\left\| \sum_{s=p}^{\infty} \frac{\varepsilon^s}{s!} f_{(s)} \right\|_{D-(\delta,\xi)} \leq \|f\|_D \frac{\left(\frac{\varepsilon}{\varepsilon^*_D}\right)^p}{1-\frac{\varepsilon}{\varepsilon^*_D}}$$

Changing now δ,ξ into $4\delta/5$, $4\xi/5$ all the assertions follow. ∎

1.4.C. Estimates of Lie derivatives

Estimates of Lie series are in a large extent estimates of Lie derivatives or, in the canonical case (to which we are interested), of Poisson Brackets. These estimates can be performed in different ways. The best one, at least for our future purposes, is that one which has been used in the proofs of the previous propositions. The idea is the following: being $L_{\mathcal{X}} f = \frac{d}{dt}(f\circ\phi^{\mathcal{X}}_t)_{t=0}$, if f and $\mathcal{X}$ are analytic then $L_{\mathcal{X}} f$ can be estimated by applying the Cauchy inequality (1) to the analytic function $t\mapsto f\circ\phi^{\mathcal{X}}_t$. In this way one finds the following result:

Lemma 1.2 Let f and $\mathcal{X}$ be analytic functions on a domain $D = D_{\rho,\sigma}(B)$, with $B\subset R^n$ an open set and $\rho,\sigma>0$. Then for any $s\geq 0$, $L^s_{\mathcal{X}} f$ is analytic on D and, for all $0<\delta<\rho$, $0<\xi<\sigma$, it satisfies

$$(15)\qquad \frac{1}{s!}\,\| L^s_{\mathcal{X}} f \|_{D-(\delta,\xi)} \leq \left(\frac{4}{\delta\xi} \|\mathcal{X}\|_D \right)^s \|f\|_D$$

Proof The analyticity of $L^s_{\mathcal{X}} f$ on D is obvious, since it is a linear combination of products of derivatives of f and $\mathcal{X}$. The estimate (15) has yet been proved in the course of the proof of proposition 1.1. ∎

To estimate multiple Poisson brackets with different Hamilton functions, say $L^{\nu_p}_{\mathcal{X}_p} \ldots L^{\nu_1}_{\mathcal{X}_1} f$, one can apply p times, in cascade, the lemma 1.2. Let us use at each step the estimate $\| L^s_{\mathcal{X}} f \|_{D-(\delta,\xi)} \leq s!(4/\delta\xi)^s \; \|\mathcal{X}_s\|^s_D \, \|f\|_D$, which is slightly worse, but easier to manage, than (15). So, for instance,

$$\| L^{\nu_2}_{\mathcal{X}_2} L^{\nu_1}_{\mathcal{X}_1} f \|_{D-(\delta_1+\delta_2,\,\xi_1+\xi_2)} \leq \nu_2!\, \left(\frac{4}{\delta_2\xi_2}\right)^{\nu_2} \|\mathcal{X}_2\|^{\nu_2}_{D-(\delta_1,\xi_1)} \, \| L^{\nu_1}_{\mathcal{X}_1} f \|_{D-(\delta_1,\xi_1)}$$

$$\leq \nu_1!\,\nu_2!\, \left(\frac{4}{\delta_1\xi_1}\right)^{\nu_1} \left(\frac{4}{\delta_2\xi_2}\right)^{\nu_2} \|\mathcal{X}_2\|^{\nu_2}_{D-(\delta_1,\xi_1)} \, \|\mathcal{X}_1\|^{\nu_1}_{D} \, \|f\|_D$$

A trivial iteration then gives, for all $p \geq 1$, $\nu_1, \ldots, \nu_p \geq 0$ and all (sufficiently small) $\delta_1, \xi_1, \ldots, \delta_p, \xi_p$:

$$(16)\qquad \left\| \frac{L^{\nu_p}_{\mathcal{X}_p} \cdots L^{\nu_1}_{\mathcal{X}_1}}{\nu_p! \cdots \nu_1!} \right\|_{D-\sum_{i=1}^{p}(\delta_i,\xi_i)} \leq \|f\|_D \prod_{s=1}^{p} \left[\frac{4}{\delta_s\xi_s} \|\mathcal{X}_s\|_{D-\sum_{j=1}^{s-1}(\delta_j,\xi_j)} \right]^{\nu_s}$$

(with $\delta_0 = \xi_0 = 0$). Let us note that there exist some variants to this procedure (for instance, one could disregard the fact that some equal $L_{\mathcal{X}_j}$ are groupped toghether and perform $|\vartheta| = \sum_{j=1}^{p} \vartheta_j$ successive estimates of simple Poisson brackets). The way just described is especially suitable for the applications to products of Lie series.

The simplest possibility is of taking all the δ_j, ξ_j equal, i.e. $\delta_j = \delta/p$, $\xi_j = \xi/p$ for some $\delta, \xi > 0$. When all the $\mathcal{X}_1, \dots, \mathcal{X}_p$ and f are analytic on D it is obtained in this way (with $|\vartheta| = \sum_{j=1}^{p} \vartheta_j$ and $L_j = L_{\mathcal{X}_j}$):

$$(17) \qquad \left\| \frac{L_p^{\vartheta_p} \cdots L_1^{\vartheta_1}}{\vartheta_p! \dots \vartheta_1!} f \right\|_{D-(\delta,\xi)} \leq \|f\|_D \left(\frac{4p^2}{\delta\xi} \right)^{|\vartheta|} \prod_{j=1}^{p} \|\mathcal{X}_j\|_D^{\vartheta_j}$$

However, it is not necessary that all the $\mathcal{X}_1, \dots, \mathcal{X}_p$ are defined in all of D, being sufficient that they be defined on suitable subsets. Coupling this fact with a slight modification of the previous procedure the following estimate, which will be used in appendix A, is easily established:

Lemma 1.3 Let $B \subset R^n$ be an open set and ρ, σ positive numbers; denote $D = D_{\rho,\sigma}(B)$ and consider positive numbers $\delta < \rho$, $\xi < \sigma$. Assume that, for each j=1,...,p, a function $\mathcal{X}_j$ analytic on $D'_j = D - \left(\frac{j-1}{p} + \frac{1}{2p} \right)(\delta, \xi)$ is given. Then, for any function f analytic on D and any multiindex $\vartheta = (\vartheta_1, \dots, \vartheta_p)$ it is

$$(18) \qquad \left\| \frac{L_p^{\vartheta_p} \cdots L_1^{\vartheta_1}}{\vartheta_p! \dots \vartheta_1!} f \right\|_{D-(\delta,\xi)} \leq \left(\frac{16p^2}{\delta\xi} \right)^{|\vartheta|} \prod_{j=1}^{p} \|\mathcal{X}_j\|_{D'_j}^{\vartheta_j}$$

Remarks (i) It is possible to perform estimates of Poisson brackets of analytic functions also withouth referring to the analyticity of the flow. The most intuitive idea is based on the explicity expression $L_{\mathcal{X}} f = \sum_{i=1}^{n} \left[\frac{\partial \mathcal{X}}{\partial p_i} \frac{\partial f}{\partial q_i} - \frac{\partial \mathcal{X}}{\partial q_i} \frac{\partial f}{\partial p_i} \right]$; when $\mathcal{X}$ and f are analytic on some set $D = D_{\varrho,\sigma}(B)$, the use of Cauchy inequalities (5) gives

$$\| L_{\mathcal{X}} f \|_{D-(\delta,\xi)} \leq \frac{2n}{\delta \xi} \| f \|_D \, \| \mathcal{X} \|_D$$

This kind of estimates has the disadvantage of introducing the dimension n of the space as a factor; if a multiple Poisson brackets is estimated along these lines, say $\| L_{\mathcal{X}}^s f \|$, factors of the order of n^s are introduced.

(ii) There exists a better possibility which, moreover, in the case of simple Poisson brackets gives a result better than (15). It is based on the idea of considering, at each point $z = (p,q) \in D-(\delta,\xi)$, only the linearization $t \mapsto z + tX(z)$ of the flow $t \mapsto \phi_t^X z$ (where $X = (-\frac{\partial \mathcal{X}}{\partial q}, \frac{\partial \mathcal{X}}{\partial p})$) and noting that $L_{\mathcal{X}} f(z) = \frac{d}{dt}[f(z+tX(z))]_{t=0}$. Assume $\mathcal{X}$ and f are analytic on D. Since for each $z \in D-(\delta,\xi)$ it is $\| X(z) \| = \max_i (| \frac{\partial \mathcal{X}}{\partial p_i}(z) | , | \frac{\partial \mathcal{X}}{\partial q_i}(z) |) \leq \| \mathcal{X} \|_D \cdot \max(1/\delta, 1/\xi)$, it follows that $t \mapsto z+tX(z)$ is an analytic mapping of $\{ |t| < \delta\xi / \| \mathcal{X} \|_D \}$ into D and then, applying Cauchy inequality to $t \mapsto f(z+tX(z))$, one finds

$$\| L_{\mathcal{X}} f \|_{D-(\delta,\xi)} \leq \frac{1}{\delta \xi} \| \mathcal{X} \|_D \, \| f \|_D$$

However, when this procedure is used to estimate multiple Poisson brackets there arise numerical factors which are no better than the $s!4^s$ of (15).

CHAPTER 2

HAMILTONIAN PERTURBATION THEORY AND NEKHOROSHEV-LIKE ESTIMATES

2.1 INTRODUCTION

Hamiltonian perturbation theory deals with the study of nearly integrable Hamiltonian systems, namely systems described by a Hamiltonian of the type

$$(1) \qquad H(p,q,\varepsilon) = h(p) + \varepsilon f(p,q,\varepsilon)$$

where $(p,q) \in R^n \times T^n$ are action-angle variables for the unperturbed, integrable system described by h, ε is a "small" parameter and the perturbation f is assumed to be regular in ε. For $\varepsilon=0$ the motion is trivial, and in particular the actions $p_1,\ldots,p_n$ are constant. For $\varepsilon\neq 0$, instead, the only result which can be easily stated by inspecting the Hamiltonian (1) is the very poor apriori estimate $|p_j(t)-p_j(0)| = O(\varepsilon t)$ $(j=1,\ldots,n)$. The purpose of perturbation theory is essentially that of going beyond this elementary result, by studying the asymptotic behaviour of the perturbed system, namely on times of the order of some positive power of $1/\varepsilon$ (or also longer) and, in particular, by obtaining estimates on the variations of the actions on these time intervals.

Basically, the perturbative approach goes as follows. With a canonical transformation one tries to annihilate, or at least to reduce as more as possible, the angle dependence of the Hamiltonian at first order in ε; then, with a second canonical transformation, one does the same at order ε^2, and so on. After this program, which is a sort of normalization procedure, has been successfully performed r times, one has an Hamiltonian canonically coniugated to the original one (at least on some subset of the original phase space) which is of the type

$$(2) \qquad H'(p,q,\varepsilon) = h(p) + \sum_{s=1}^{r} \varepsilon^s Z_s(p,q) + \varepsilon^{r+1} R_{r+1}(p,q,\varepsilon)$$

where the functions Z_s depend only on some of the angles, say on $0 \le m < n$ of them, or more generally on $m < n$ linear combinations of the angles; no condition is instead imposed on the remainder R_{r+1} (except to be regular for $\varepsilon \to 0$). Obviously, to arrive at (2) one could employ, instead than a sequence of canonical transformations (realized, for instance, with a composition of Lie transforms), a single one (a Deprit's direct or inverse transform with a generating function which is a polynomial of degree r in ε, for instance).

In particular, if all the angles can be dropped out of $Z_1, \ldots, Z_r$ in (2), then the variation of the momenta $p_1, \ldots, p_n$ is of order ε on times of order $(1/\varepsilon)^r$; more precisely, if P(t) is a solution of the Hamilton equations for (2), then

$$\sup_{|t| \le (\frac{1}{\varepsilon})^r} \| P(t) - P(0) \| \le \varepsilon \left\| \frac{\partial R_{r+1}}{\partial \varepsilon} \right\|$$

(with suitable norms). On the other hand, for the solutions $p(t)$ of (1) one has $\|p(t)-p(0)\| \leq \|p(t)-P(t)\| + \|P(t)-P(0)\| + \|P(0)-p(0)\|$, where the first and last terms can be controlled with some norm of the canonical transformation. Thus, if the canonical transformation is ε-close to the identity, an overall estimate

$$\|p(t)-p(0)\| \leq \varepsilon \cdot \text{const} \qquad \text{for all} \quad |t| \leq \left(\frac{1}{\varepsilon}\right)^r$$

is obtained, i.e. there exists a "stability" time $(1/\varepsilon)^r$ for the motions of the perturbed systems (1).

As mentioned in the Introduction, Nekhoroshev has succeeded in obtainin a stability time which is exponentially long in $1/\varepsilon$ and then longer than any power of $1/\varepsilon$. Actually, the classical setting just outlined quite naturally leads to Nekhoroshev's exponential estimates: if one can show that, for each (sufficiently small) ε, this normalization procedure can be performed a number of times which is a negative power of ε, say $r \sim \varepsilon^{-b}$ for some $b>0$, a stability time of the order of $\varepsilon^{-(1/\varepsilon)^b}$ or $e^{-(1/\varepsilon)^b}$ is then obtained. Let us stress that, for a reason explained in section 3, on this time scale one obtains a bound $\|p(t)-p(0)\| \leq \text{const} \cdot \varepsilon^a$ for some $0<a<1$.

This basic idea is well illustrated by the simple case of a linear unperturbed Hamiltonian $h = \omega \cdot p$, namely when $H = h+\varepsilon f$ describes a system of (weakly) coupled harmonic oscillators, and, moreover, a strong nonresonance condition is satisfied.

Indeed, it has been proven in [11] that, when the constant frequency ω satisfies a diofantine condition, exponential estimates for the stability time are obtained. We will give a full treatment of this case (section 3 and appendix A), employing a product of Lie series to generate the "normalizing" canonical transformation.

In order to treat the nonconstant frequency case, which is the content of Nekhoroshev's original work [38,39], one needs some modifications of this simple idea. Actually, when performing the normalization procedure one is faced with the well known difficulties due to the presence of resonances. Resonances manifest themselves analitycally in the so called "small denominators", which can make divergent the formal series defining the normalizing canonical transformations. A very remarkable part of Nekhoroshev's work is precisely the overcoming of the small denominators difficulty. Basically, this aim is achieved by decomposing the phase space in different regions ("blocks"), depending on the local nonresonance conditions, and performing a different normalization in each of them; moreover, the so called "ultraviolet cutoff" technique, introduced by Arnold in his proof of KAM theorem, must be employed. Furthermore, in this way one does not achieve the complete elimination of all the angles from the normalized Hamiltonian in each block, and some further work is needed to prove an ε-confinement of the actions. We will give only a sketch of the whole procedure (section 2), while treating in detail (section 4 and appendix B) the normalization inside a given block. The result so obtained will then be used, to provide a concrete example of the use of the cutoff technique, to re-study the case of a system of diofantine harmonic oscillators.

2.2 NORMALIZATION

The aim of this section is to provide a general introduction to the normalization procedure and to Nekhoroshev's exponential estimates. The presentation of the pertubative techniques will be necessarily very short and focalized to the attainment of these estimates; for broader introduction and/or a deeper analysis of the effects of the resonances see for instance [5,10,11,19,23].

Within this section we will consider in detail only the first order normalization, which is fairly sufficient to focalize the main features of the subject; the higher order normalization will be fully considered in the following sections. Furthermore, we assume for simplicity that the perturbation is linear in the parameter ε, thus dealing with a Hamiltonian

$$(1) \qquad H_\varepsilon(p,q) = h(p) + \varepsilon f(p,q).$$

which will be assumed to be regular on some set $B \times T^n$, $B \subset R^n$, or also, if needed, analytic on a complex set $D = D_{\varrho,\sigma}(B)$ defined as in section 1.4. We always denote by $\omega(p) = \frac{\partial h}{\partial p}(p)$ the frequency vector of the unperturbed system. Let us mention that we will be mainly concerned with the two extreme cases of a linear unperturbed Hamiltonian $h = \omega \cdot p$, ω =const, and of a nondegenerate one, i.e.

$$(2) \qquad \det\left(\frac{\partial^2 h}{\partial p \partial p}(p)\right) \neq 0 \qquad \text{at all } p.$$

2.2.A Preliminaries

We first recall some basic definitions about resonances and some elementary properties of Fourier series.

Resonances A point p is said to be a resonant point iff there exists an integer non null vector $k=(k_1,\ldots,k_n) \in \mathbb{Z}^n$, such that $k\cdot\omega(p)=0$. The vector $\omega(p)$ is then said to be resonant with k; $|k| = \sum_{j=1}^{n} |k_j|$ is the order of the resonance. If $\omega(p)$ resonates with m linearly independent (over the rationals) integer vectors $k^{(1)},\ldots,k^{(m)}$ then we say that a resonance of multiplicity m takes place.

Lattices To characterize resonances it is useful the notion of resonant lattice. Let us denote by Q the set of all the rational numbers. Consider the vector space Q^n (over Q) and regard $\mathbb{Z}^n$ as a subset of Q^n. Then we define as a m-dimensional ($0\leq m \leq n$) lattice $\mathcal{M}$ of $\mathbb{Z}^n$ the intersection of $\mathbb{Z}^n$ with a m-dimensional linear subspace of Q^n. An integer basis of a m-dimensional lattice $\mathcal{M}$ is any set of m linearly independent (over Q) vectors $k^{(1)},\ldots,k^{(m)}$ which are in $\mathcal{M}$ and such that any $k \in \mathcal{M}$ can be written $k = \sum_j c_j k^{(j)}$ with rational coefficients c_j.

The set of all $k \in \mathbb{Z}^n$ such that $k \cdot \omega(p) = 0$ is obviously a lattice of dimension equal to the multiplicity of the resonance at p. It will be called the maximal resonant lattice at p; any lattice contained in it is also a resonant lattice at p. Note that if the zero-dimensional lattice $\mathcal{M} = \{0\}$ is also taken into consideration, to each point p one can associate a resonant lattice.

Resonant surfaces To each lattice $\mathcal{M}$ we associate its resonant surface

$$\Sigma_{\mathcal{M}} = \{p \in B: \omega(p) \cdot k = 0 \text{ for all } k \in \mathcal{M}\}.$$

When h satisfies the nondegeneracy condition (2) each $\Sigma_{\mathcal{M}}$ is either empty or a smooth submanifold of B of dimension n−m, where $m = \dim \mathcal{M}$. In particular, $\Sigma_{\{0\}} = B$ while $\Sigma_{\mathbb{Z}^n}$ is made only of those isolated points at which ω vanishes.

Let us stress that, when h is nondegenerate, the set of points having a resonant lattice of any dimension m, $0 \leq m < n$, as well as the set of the nonresonant points, is dense in B. This immediately follows from the fact that, under (2), $\omega : B \to R^n$ is a local diffeomorphism and from the density in R of rationals and irrationals. This means that, for each $0 \leq m < n$, $\bigcup_{\substack{\mathcal{M} \\ \dim \mathcal{M} = m}} \Sigma_{\mathcal{M}} \setminus \bigcup_{\substack{\mathcal{M}' \\ \dim \mathcal{M}' > m}} \Sigma_{\mathcal{M}'}$ densely fills B.

Fourier series For any function f on $B\times T^n$ (or $D_{\rho,\sigma}(B)$) we denote by $\tilde{f}_K$ its Fourier components, so that

$$f(p,q) = \sum_{K\in\mathbb{Z}^n} \tilde{f}_K(p)\, e^{iK\cdot q}$$

For any lattice $\mathcal{M}$ we denote by $\Pi_{\mathcal{M}}$ the projector on the subspace of functions having nonzero Fourier components only on $\mathcal{M}$, i.e.

$$[\Pi_{\mathcal{M}} f](p,q) = \sum_{K\in\mathcal{M}} \tilde{f}_K(p)\, e^{iK\cdot q}$$

Note that when $\mathcal{M} = \{0\}$, $[\Pi_{\{0\}} f](p) = \tilde{f}_0(p) = (2\pi)^{-n}\int_{T^n} f(p,q)\,dq$ is the average of $f(p,\cdot)$ over the torus $\{p\}\times T^n$. It turns out that, for any lattice $\mathcal{M}$ of dimension m, $\Pi_{\mathcal{M}}$ is an averaging operator over some (n-m)-dimensional subtorus of T^n. To achieve this result, one first shows that, for each m-dimensional lattice $\mathcal{M}$ there exists an integer matrix J with determinant 1, such that

$$J\mathcal{M} = \left\{ k=(k_1,\dots,k_n)\in\mathbb{Z}^n :\ k_{m+1}=\dots=k_n=0 \right\}$$

(for the proof, see [11]). It follows that J is an automorphism of T^n and it is then possible to define new angular coordinates $(\sigma_1,\dots,\sigma_m,\varphi_{m+1},\dots,\varphi_n) = \tilde{J}^{-1}(q_1,\dots,q_n)$ on T^n. Thus a factorization $T^n \approx T^m\times T^{n-m}$ is obtained, with $(\sigma_1,\dots,\sigma_m)\in T^m$ and $(\varphi_{m+1},\dots,\varphi_n)\in T^{n-m}$. After this, it is easy to verify (see [11]) that

$$[\Pi_K f](p,q) = \frac{1}{(2\pi)^{n-m}} \int_{T^{n-m}} f(p, \tilde{J}(\sigma,\varphi))\, d\varphi$$

The Fourier components of complex analytic functions have the property, which plays an essential role in perturbation theory, of decreasing exponentially with $|k| = \sum_{j=1}^{n} |k_j|$. The same happens for the tail of the series, too: to make this precise, consider any positive integer K, called a cutoff, and perform the decomposition $f = f^{\le K} + f^{>K}$, with

$$(3) \qquad f^{>K}(p,q) = \sum_{|k|>K} \tilde{f}_k(p)\, e^{ik\cdot q} ;$$

then one has the following

Lemma 2.1 Let f be any analytic function on a set $D = D_{\rho,\sigma}(B)$, with $B \subset R^n$ and $\rho, \sigma > 0$. Then, for each $k \in \mathbb{Z}^n$, $\tilde{f}_k$ is analytic on $D_\rho(B)$ and satisfies

$$(4) \qquad \|\tilde{f}_k\|_D \le \|f\|_D\, e^{-|k|\sigma}$$

(where $\|\tilde{f}_k\|_D = \|\tilde{f}_k\|_{D_\rho(B)}$). Moreover, for any $K \ge 0$ and any $0 < \xi < \min(\sigma, 1)$ one has

$$(5) \qquad \|f^{>K}\|_{D-(0,\xi)} \le \|f\|_D \left(\frac{8}{\xi}\right)^n e^{-K\frac{\xi}{2}}$$

Proof The analyticity of $\tilde{f}_k$ on $D_\rho(B)$ follows from that of f and its definition

$$\tilde{f}_k(p) = (2\pi)^{-n} \int_{T^n} f(p,q)\, e^{-ik\cdot q}\, dq$$

By suitably shifting the path of integration one can also write, for any $0<\delta<\sigma$:

$$\tilde{f}_k(p) = (2\pi)^{-n} \int_{T^n} f(p, q - i\frac{k}{|k|}(\sigma-\xi))\; e^{-ik\cdot\left[q - i\frac{k}{|k|}(\sigma-\xi)\right]}\, dq$$

Thus, denoting $\mathcal{S}_{\sigma-\xi} = \{ q \in C^n : \|\mathrm{Im}(q)\| < \sigma-\xi \}$, we have

$$|\tilde{f}_k(p)| \le \sup_{q\in\mathcal{S}_{\sigma-\xi}} |f(p,q)|\; (2\pi)^{-n}\int_{T^n} \left| e^{ik\cdot q}\; e^{-|k|(\sigma-\xi)} \right| dq \le$$

$$\le \sup_{q\in\mathcal{S}_{\sigma-\xi}} |f(p,q)|\; e^{-|k|(\sigma-\xi)}$$

So $\|\tilde{f}_k\|_D \le \|f\|_D \exp[-|k|(\sigma-\xi)]$ for all $0<\xi<\sigma$ and (2) follows.

Now, being $|ik\cdot q| \le |k|\,\|\mathrm{Im}(q)\| \le |k|(\sigma-\xi)$ for all $q\in\mathcal{S}_{\sigma-\xi}$, one has

$$\| f^{>K} \|_{D-(0,\xi)} \le \sum_{|k|>K} \|\tilde{f}_k\|_D \quad e^{|k|(\sigma-\xi)}$$

and it then follows from (4) that

$$\|f^{>K}\|_{D-(0,\xi)} \le \|f\|_D \sum_{|k|>K} e^{-|k|\xi} \le \|f\|_D\, e^{-K\xi/2} \sum_{|k|>K} e^{-|k|\xi/2}$$

The last sum is easily estimated: using $|k| = \sum_{j=1}^{n} |k_j|$ one has

$$\sum_{|k|>K} e^{-|k|\xi/2} \le \sum_{k\in\mathbb{Z}^n} e^{-|k|\xi/2} = \left[\sum_{l=-\infty}^{+\infty} e^{-|l|\xi/2} \right]^n =$$

$$= \left[-1 + 2\sum_{l=0}^{\infty} e^{-l\xi/2} \right]^n = \left[\frac{1+e^{-\xi/2}}{1-e^{-\xi/2}} \right]^n \le \left(\frac{8}{\xi}\right)^n$$

where the last inequality follows from $\exp(-\xi/2)] \ge 1-\xi/4$ for all $\xi \le 2$. ∎

2.2.B Normalization

The general features of the (first order) normalization procedure emerge by looking at the first order effects of an ε-close to the identity canonical transformation on a nearly integrable Hamiltonian $H_\varepsilon = h+\varepsilon f$. For definitness, let us generate the transformation with a simple Lie transform ϕ^{χ}.

As seen in section 1.3 we have

$$(6) \qquad (\phi^{\chi}_{\varepsilon})^{*}[h+\varepsilon f] = h + \varepsilon[L_{\chi}h + f] + \sum_{s\geq 2} \frac{\varepsilon^{s}}{(s-1)!} L_{\chi}^{s-1} [\tfrac{1}{s} L_{\chi}h + f]$$

Let us write this expression in the form

$$(7) \qquad (\phi^{\chi}_{\varepsilon})^{*}[h+\varepsilon f] = h + \varepsilon Z + \varepsilon^{2} R$$

Then the generating function χ and the new first order term Z must be solutions of the equation

$$(8) \qquad L_h \chi = f - Z$$

(where $L_h \chi = -L_{\chi} h$ has been used). As it will be seen later, when performing the higher order normalization an equation of exactly this type has to be solved at any order.

The properties of the (possible) solutions χ, Z of (8) are easily exploited by expanding all the functions in Fourier series ; with the notation previously introduced, (8) becomes, at each point $p \in B$ and for each $k \in \mathbb{Z}^{n}$:

$$(9) \qquad i\,\omega(p)\cdot k\ \widetilde{\mathcal{X}}_k(p) = \widetilde{f}_k(p) - \widetilde{Z}_k(p)$$

Let us denote by $\mathcal{M}_p$ the (maximal) resonant lattice of $\omega(p)$; then equation (9) implies that, for each $p \in B$:

(a) it must be $\widetilde{f}_k(p) = \widetilde{Z}_k(p)$ for all $k \in \mathcal{M}_p$, i.e.

$$\Pi_{\mathcal{M}_p} f(p,\cdot) = \Pi_{\mathcal{M}_p} Z(p,\cdot),$$

while no condition is imposed on $\widetilde{Z}_k(p)$ for $k \notin \mathcal{M}_p$;

(b) it must be, for all $k \notin \mathcal{M}_p$,

$$\widetilde{\mathcal{X}}_k(p) = \frac{\widetilde{f}_k(p) - \widetilde{Z}_k(p)}{i\,k\cdot\omega(p)}$$

while no condition is imposed on $\widetilde{\mathcal{X}}_k(p)$ for $k \in \mathcal{M}_p$.

Note that (a) means that, at each point $p \in B$, the "resonant" harmonics of the perturbation (i.e. those f such that $\widetilde{f}_k(p)=0$ and $\omega(p)\cdot k=0$) can not be killed. Moreover,to save the continuity of each $\widetilde{Z}_k$, and thus of Z, a Fourier component which is resonant at a point p has to survive in Z also in some neighbourhood of p. It can be expected that, in generic situations in which the resonances change from point to point, this fact implies that a so large number of Fourier components of the perturbation has to survive in Z that the normalization may become completely useless. We will return on this later.

The goal of the normalization procedure is that of obtaining, at least in some open subset $B'\times T^n$ of the whole phase space, a normal form (7) with

$$(10) \qquad Z = \Pi_{\mathcal{M}} f \qquad \text{on all of } B'\times T^n$$

where $\mathcal{M}$ is some lattice of dimension $m<n$, so that the new Hamiltonian in $B'\times T^n$ is

$$(11) \qquad H\circ\phi_\varepsilon^{\mathcal{X}} = h + \varepsilon\,\Pi_{\mathcal{M}} f + \varepsilon^2 R$$

Let us call this expression a first order normal form of H_ε adapted to the lattice $\mathcal{M}$. Its usefulness is in the fact that the system described by (11) possesses n−m independent "approximate" (at first order) constants of the motion which are linear combinations of actions; indeed, for any $v\in R^n$ orthogonal to $\mathcal{M}$, one has

$$\frac{d}{dt}(v\cdot p) = \left\{v\cdot p, H\circ\phi_\varepsilon^{\mathcal{X}}\right\} = \varepsilon^2 \left(v\cdot\frac{\partial R}{\partial q}\right)$$

so that $v\cdot p$ varies of quantities of order ε over times $t\sim 1/\varepsilon$.

The geometrical meaning of this fact is the following. Let us denote $\Lambda_{\mathcal{M}}(p)$ the m-dimensional ($m=\dim\mathcal{M}$) affine subspace of R^n through the point p and parallel to $\mathcal{M}$. Consider any motion $(P(t),Q(t))$ corresponding to the normalized Hamiltonian (11), which starts from a point $(P(0),Q(0))\in B'\times T^n$ (and remains there for $|t|\le 1/\varepsilon$). Then, for all $|t|\le 1/\varepsilon$, $P(t)$ is ε-closed to $\Lambda_{\mathcal{M}}(P(0))$: up to small ($\sim\varepsilon$) deviations in transversal directions, the motion of the actions in the normalized system is forced to take place on a m-dimensional hyperplane (the "plane of fast drift").

The same conclusions hold up to times $|t|\sim(1/\varepsilon)^r$ when H can be putted in a r-th order normal form adapted to $\mathcal{M}$, i.e. when there exists a canonical transformation ϕ_ε such that

$$H_\varepsilon \circ \phi_\varepsilon = h + \sum_{s=1}^{r} \varepsilon^s Z_s + O(\varepsilon^{r+1})$$

with $\Pi_{\mathcal{M}} Z_s = Z_s$ for all s=1,...,r.

However, as known from Poincare's times, a (nontrivial) normal form adapted to a lattice does not exists in generic situations. For instance, if

(i) the unperturbed Hamiltonian h satisfies the nondegeneracy condition (2) and

(ii) the perturbation f has "sufficiently many" Fourier components, in the sense that, at each point $p \in B$ and for each lattice $\mathcal{M}$, $\left(\Pi_{\mathcal{M}} f\right)(p,\cdot)$ is non-zero,

then there exists no regular near to the identity canonical transformation $\phi : I \times B' \times T^n \to B \times T^n$, where I is some interval containing 0 and B' some open nonempty subset of B, such that $H_\varepsilon \circ \phi_\varepsilon$ is a normal form adapted to a lattice $\mathcal{M}$ of dimension m<n. In other words, under these hypotheses, the unique normal form of H_ε adapted to a lattice is H_ε itself.

The proof of this fact is very simple. First, note that the generating function of the canonical transformation which gives H the normal form (11) should be given by

$$(10') \qquad \mathcal{X}(p,q) = \sum_{\kappa \notin \mathcal{M}} \frac{\tilde{f}_\kappa(p)}{i\kappa\cdot\omega(p)} e^{i\kappa\cdot q}$$

(actually, to this $\mathcal{X}$ one could add any function Ξ such that $\Pi_{\mathcal{M}}\Xi = \Xi$; however, its presence is rather inessential and we will take it, as usual, equal to zero). Because of condition (ii) and to save the continuity of $\mathcal{X}$, we have to include in $\mathcal{M}$ all the maximal resonant lattices $\mathcal{M}_p$, $p \in B'$, i.e. it must be $\mathcal{M} \supset \bigcup_{p\in B'} \mathcal{M}_p$. But, by (i), $\omega(B') = \bigcup_{p\in B'} \omega(p)$ is open and non-empty in R^n, so that it is not contained in any subspace of R^n of positive codimension; clearly, the same is true also for $\bigcup_{p\in B'} \mathcal{M}_p$.

Remarks (i) This impossibility result is essentially part of a well known theorem of Poincare on the nonexistence of integrals of motion in nearly integrable Hamiltonian systems, which has been felt for a long time as ruling out the very possibility of a successful perturbation theory; for a deeper discussion see [10].

(ii) The condition that the normalizing canonical transformation is realized as a simple Lie transform is completely irrelevant. However, what is essential for this conclusion is the regularity in ε, namely that the normalization is performed with a family $\phi : I\times B'\times T^n \to B\times T^n$ of regular canonical transformations, where I is an interval containing zero. Actually, in Nekhoroshev's overcoming of this "impossibility", one looses the regularity of the canonical transformation in ε

for $\varepsilon = 0$; this is not dramatic, since the transformation do exists for each $\varepsilon \neq 0$, while one clearly does not need it for $\varepsilon = 0$; for a discussion of this fact see [10].

The foregoing conclusion on the unfeasibility of the normalization does not hold in special cases in which at least one of the two conditions (i) and (ii) is violated. There are two especially important cases. The first one is when h is linear in p, so that the frequency $\omega = \frac{\partial h}{\partial p}$ is constant and a single resonant lattice appears; this case is briefly considered in the next subsection.

The second one is that of a perturbation having only a finite number of Fourier components, say $\tilde{f}_K = 0$ for all $|k| > K$. In this case one has to take into account only the resonances of order $|k| \leq K$, which are finitely many, and the corresponding resonant surfaces $\Sigma_{\mathcal{M}}$ ($\mathcal{M} \neq \{0\}$) no longer fill densely B. One can then decompose B into subsets $B_{\mathcal{M}}$, called "blocks", each of which contains (part of) only one of these resonant surfaces $\Sigma_{\mathcal{M}}$; in each block $B_{\mathcal{M}}$ it is possible to give H_ε a normal form adapted to the lattice $\mathcal{M}$. In this way, which has fully developed by Nekhoroshev in the proof of his theorem, the "impossibility" of the normalization is confined to the block $B_{\mathbb{Z}^n}$, which consists of a small neighbourhood of the possible point at which ω vanishes, while in all the other blocks a normal form adapted to a lattice of dimension lesser than n can be obtained. The importance of this special case is in the fact that it is always possible to reduce oneself to it by using a cutoff. We will return on this in subsection D, where some details on the decomposition into blocks and on its use are also given.

2.2.C The fixed frequency case

Assume that $\omega = \frac{\partial h}{\partial p}$ is constant and denote by $\mathcal{M}_\omega$ its resonant lattice. According to the previous discussion, the best one can do is to look for a solution (10) with $\mathcal{M} = \mathcal{M}_\omega$, namely $Z = \Pi_{\mathcal{M}_\omega} f$ and

$$(12) \qquad \mathcal{X}(p,q) = \sum_{k \notin \mathcal{M}_\omega} \frac{\tilde{f}_k(p)}{i k \cdot \omega} e^{i k \cdot q}$$

Here Z is well defined (analytic, if f is) but, concerning $\mathcal{X}$, the "small denominators" difficulty appears: although all the zeros of $\omega \cdot k$ have been excluded from (12), the quantity $|\omega \cdot k|$ becomes arbitrarily small for suitable $k \in \mathbb{Z}^n$ (near to the subspace of R^n orthogonal to $\mathcal{M}_\omega$), because of the very possibility of approximatinge irrational numbers with rational ones. Thus, to assure the convergence of the series, and then the very existence of the canonical transformation, some additional condition needs to be satisfied.

An obvious possibility is that f has only finitely many Fourier components, say $\tilde{f}_k = 0$ for $|k| > K$, since then the sum in (12) is finite . This is the case of the so called Birkhoff's normalization. We do not consider it explicitely, as it can be recovered as a special case of the treatment of the next subsection and of section 2.4.

There is another, very important, condition which permits the overcoming of the small denominator difficulty. Let us first discuss it in the nonresonant case, namely when $\mathcal{M}_\omega = \{0\}$, so that $Z = \tilde{f}_0$ and

$$(13) \qquad \mathcal{X}(p,q) = \sum_{k \neq 0} \frac{\tilde{f}_k(p)}{i k \cdot \omega} e^{i k \cdot q}$$

It is easy to recognize that in the (complex) analytic case, because of the exponential decay of the $\tilde{f}_k$, a sufficient condition for the convergence of this series is that $|\omega \cdot k|$ does not decrease more rapidly than some power of $|k|$ for $|k| \to \infty$, namely that ω satisfies a "diofantine" condition

$$(14) \qquad |\omega \cdot k| \geq \gamma |k|^{-a} \qquad \text{for all } k \in \mathbb{Z}^n,\ k \neq 0$$

for some positive numbers γ and a. For the details, see the proposition A.1 in appendix A.

The importance of the diofantine condition (14) is that, provided a is sufficiently large (for instance $a=n$), the set of those vectors $\omega \in R^n$ which satisfy it for at least a $\gamma = \gamma(\omega)$ is of Lebesgue measure one [2,6,37]. Let us mention that, however, for any fixed γ and a, the set of $\omega \in R^n$ which do not satisfy (14) is open and dense in R^n.

The next section is devoted to a detailed treatment of the diofantine case, assuming for definitness a =n. Let us anticipate that, under this condition, the normalization can be performed up to any (finite) order and the ε-confinment of the actions is then achieved.

The case of a resonant frequency, with $\mathcal{M}_\omega \neq \{0\}$, can be treated in a similar way if ω satisfies a diofantine condition outside $\mathcal{M}_\omega$, i.e. $|\omega \cdot k| \geq \gamma |k|^{-a}$ for all $k \notin \mathcal{M}_\omega$. Under this condition, a normal form adapted to $\mathcal{M}_\omega$ can be obtained (see also [10]). As far as dim $\mathcal{M}_\omega > 0$, however, no ε-confinment of the actions is achieved since, as previously discussed, the actions remain free to move along the plane of fast drift $\Lambda_{\mathcal{M}}$.

2.2.D The nondegenerate case

We present in this last subsection the general lines of Nekhoroshev's like normalization for nonconstant frequency systems, which is based on the use of a cutoff and on the decomposition of phase space in blocks, and then indicate how this procedure can be used to obtain exponentially long stability times. It is here assumed that $H=h+\varepsilon f$ is analytic on a set $D_{\varrho,\sigma}(B)$ and that h is there nondegenerate (additional conditions, for instance convexity of h, are later needed for the stability).

Cutoff The method of the so called "ultraviolett" cutoff, introduced by Arnold in his own proof of KAM, is based on the idea that, because of the exponential decay of the Fourier components of analytic functions, the contribution to the dynamics of the high frequency part of the perturbation is so small that it can be ignored in the normalization procedure. Indeed, consider any integer $K>0$ and decompose the Hamiltonian $H=h+\varepsilon f$ as $H=H^{\leq K}+H^{>K}$, with $H^{>K}=f^{>K}$ (see (3)). Assume now that, in some complex neighbourhood $D=D_{\rho,\sigma}(B')$ of a set $B'\times T^n$, $H^{\leq K}=h+\varepsilon f^{\leq K}$ can be put, via an analytic canonical transformation $\phi_\varepsilon^{\mathcal{K}}$, in a normal form adapted to some lattice $\mathcal{M}$, i.e.

$$(\phi_\varepsilon^{\mathcal{K}})^* [h+\varepsilon f^{\leq K}] = h + \varepsilon \Pi_{\mathcal{M}}(f^{\leq K}) + \varepsilon^2 R(h+\varepsilon f^{\leq K})$$

(we will discuss below this possibility). Then, the full Hamiltonian H_ε is canonically coniugated in D to

$$(15)\quad (\phi_\varepsilon^{\mathcal{K}})^*[h+\varepsilon f] = h + \varepsilon \Pi_{\mathcal{M}}(f^{\leq K}) + \varepsilon^2 R(h+\varepsilon f^{\leq K}) + f\circ\phi_\varepsilon^{\mathcal{K}}$$

Here the remainder contains the additional term $f^{>K}\circ\phi_\varepsilon^{\mathcal{K}}$; to estimate it let us notice that, if $\phi_\varepsilon^{\mathcal{K}}$ is near to the identity, one has, on some subset $D-(\delta,2\xi)$ of D:

$$\|\varepsilon\ f^{>K}\circ\phi_\varepsilon^{\mathcal{K}}\|_{D-(\delta,2\xi)} \leq \varepsilon\|f^{>K}\|_{D-(0,\xi)} \leq \varepsilon\|f\|_D\left(\frac{\varepsilon}{\xi}\right)^n \exp(-K\xi/2)$$

where the last inequality follows from lemma 2.1. So, it is sufficient to choose the cutoff K sufficiently large, say $K\sim \ln(1/\varepsilon)$, to have that $\|\varepsilon f^{>K}\circ\phi_\varepsilon^{\mathcal{K}}\|$ is of order ε^2. A similar procedure can be used for higher order normalization (see section 2.4). Note that if K can be choosen to be a negative power of ε, say $K \sim (1/\varepsilon)^b$, then the contribution of the ultraviolet part $f^{>K}$ of the perturbation to the transformed Hamiltonian (15) is of order $\exp[-(1/\varepsilon)^b]$.

Nekhoroshev's decomposition We now consider the decomposition of phase space needed to normalize $H_\varepsilon^{\leq K} = h + \varepsilon f^{\leq K}$. Let us introduce the following terminology: a m-dimensional lattice $\mathcal{M}$ is called a K-lattice iff it has an integer basis (defined as in 2.2.A) made of vectors $k^{(1)}, \dots, k^{(m)}$ such that $|k^{(j)}| \leq K$ for all $j=1,..,m$; such a basis will be called a K-basis; by convention, also the zero-dimensional lattice $\mathcal{M} = \{0\}$ is a K-lattice. Clearly, to normalize $H_\varepsilon^{\leq K}$ it is sufficient to take into consideration only the resonant surfaces $\Sigma_{\mathcal{M}}$ corresponding to K-lattices.

Consider a sequence of positive numbers $\alpha_0, \alpha_1, \dots, \alpha_n$ such that $\alpha_n \geq 2\alpha_{n-1} \geq .. \geq 2\alpha_0$. Let us associate to each m-dimensional ($1 \leq m \leq n$) K-lattice its resonant surface $\Sigma_{\mathcal{M}}$ and, around it, a resonant zone of "thickness" α_m, defined as

$$Z_{\mathcal{M}} = \left\{ p \in B:\ |\omega(p) \cdot k^{(j)}| < \alpha_m \ \text{for at least a K-basis}\ k^{(1)}, \dots, k^{(m)} \ \text{of}\ \mathcal{M} \right\};$$

for $\mathcal{M} = \{0\}$ define $Z_{\{0\}} = B$. For each K-lattice $\mathcal{M}$ of any dimension $0 \leq m \leq n$, let us then define the $\mathcal{M}$-resonant block

$$B_{\mathcal{M}} = Z_{\mathcal{M}} \setminus \bigcup_{\substack{\mathcal{M}' \\ \dim \mathcal{M}' > \dim \mathcal{M}}} B_{\mathcal{M}'}$$

(with $B_{\mathbb{Z}^n}=Z_{\mathbb{Z}^n}$). By construction, the blocks cover B (although they are not a partition) and in each of them the nonresonance condition

$$(16) \qquad |\omega(p)\cdot k| \geq \alpha_{m+1} \quad \text{for all } k\notin \mathcal{M},\ |k|\leq K, \text{ and all } p\in B_{\mathcal{M}}$$

(with $m=\dim\mathcal{M}$) is satisfied.

Normalization inside a block In order to perform the normalization one needs to work in some complex neighbourhood of each block, say some $D = D_{\varrho_m,\sigma_m}(B_{\mathcal{M}})$ (where $m=\dim\mathcal{M}$), on which a nonresonance condition of the type (16) can be assured. Let us give some indication on this. For each $p\in D_{\varrho_m}(B_{\mathcal{M}})$ we can write

$$|\omega(p)\cdot k| \geq |\omega(p_0)\cdot k| - |[\omega(p)-\omega(p_0)]\cdot k|$$

where $p_0\in B_{\mathcal{M}}$. Now, from

$$|[\omega(p)-\omega(p_0)]\cdot k| = \left|\int_{p_0}^{p} k\cdot\frac{\partial^2 h}{\partial p\partial p}(p')\,dp'\right| \leq Kn\left\|\frac{\partial^2 h}{\partial p\partial p}\right\|_D \|p-p_0\|$$

$$\leq Kn\left\|\frac{\partial^2 h}{\partial p\partial p}\right\|_D \varrho_m$$

we get $|\omega(p)\cdot k| \geq |\omega(p_0)\cdot k| - Kn\,\varrho_m\left\|\frac{\partial^2 h}{\partial p\partial p}\right\|_D$. So, if the various parameters satisfy the condition

$$(17) \qquad Kn\,\varrho_m \left\|\frac{\partial^2 h}{\partial p\partial p}\right\|_D \leq \alpha_m$$

we then obtain, using (16) and $\alpha_{m+1}\geq 2\,\alpha_m$, that for all $p\in$

$D_{\rho_m}(B_{\mathcal{K}})$ one has

$$(18) \qquad |\omega(p)\cdot k| \geq \alpha_m \quad \text{for all } k \notin \mathcal{K}\,, \quad |k| \leq K.$$

It is easy to recognize that this nonresonance condition assures that $\mathcal{X}(p,q) = \sum_{k\notin\mathcal{K}} \tilde{f}_k(p)\cdot[ik\cdot\omega(p)]^{-1}\exp(ik\cdot q)$ is analytic on $D_{\rho_m,\sigma_m}(B_{\mathcal{K}})$ (see also appendix B) and that $H_\varepsilon^{\leq K}$ is then canonically coniugated in $B_{\mathcal{K}}\times T^n$ to

$$(19) \qquad h + \varepsilon\,\Pi_{\mathcal{K}} f^{\leq K} + \varepsilon^2 R(h+\varepsilon f)^{\leq K} + \varepsilon f^{>K}\circ\phi_\varepsilon^{\mathcal{X}}$$

The confinment of the actions As previously noticed, the usefulness of the normal form (19) is in that it implies that the images P of the actions under the canonical transformations, in each block $B_{\mathcal{K}}$, move freely along the plane of fast drift $\Lambda_{\mathcal{K}}(P(0))$, but are confined to an ε-neighbourhood of it for times $|t| \leq 1/\varepsilon$. So, to get a bound $\|P(t)-P(0)\| \leq \varepsilon$ for $|t|\leq 1/\varepsilon$ one needs some more condition which assures a confinment of the projection of the motion on $\Lambda_{\mathcal{K}}(P(0))$.

A sufficient condition for this is the convexity of h, and in this case the mechanism which gives the confinment is simply the conservation of energy [11]. Indeed, assume that $\frac{\partial^2 h}{\partial p\,\partial p}$ is everywhere positive definite. It is easy to see that this implies that

(i) each hyperplane $\Lambda_{\mathcal{K}}$ intersects transversally the corresponding resonant surface $\Sigma_{\mathcal{K}}$, and

(ii) h restricted to $\Lambda_{\mathcal{K}}$ has a minimum in the point P^* of the intersection of Λ_{μ} and Σ_{μ}, so that the projection of P(t) on the plane of fast drift $\Lambda_{\mu}(P(0))$ cannot escape from a convenient surrounding of P^*, of semidiameter approximately given by dist$(P(0),P^*)$.

If each block B_{μ} is small with ε (as can be assured by taking the α_j small with ε), the ε-confinment of the actions is achieved. Similar considerations hold for the higher order normal forms. Exponential estimates for the time of confinment are then obtained by showing that, inside each block, the normalization can be performed up to an order r and with a cutoff K which are negative powers of ε .

Greater details on the proof of Nekhoroshev theorem, as well as its precise formulation, can be found in [38,39,9,11]. We will give in the following (section 4 and appendix B) a quite detailed treatment only of the "analytic" part of the proof, namely the normalization inside a block.

2.3 EXPONENTIAL ESTIMATES IN THE FIXED FREQUENCY CASE

We consider in this section the case of a system of (weakly) coupled, strongly nonresonant harmonic oscillators. We first present in a formal way the normalization procedure up to some finite order r, employng a product of Lie series (any other method could be used). After this, we enunciate a rigorous reasult whose proof is deferred to Appendix A. Nekhoroshev-like exponential estimates are finally obtained.

Consider an Hamiltonian $H_\varepsilon = h + \varepsilon f$ and r functions $\mathcal{X}_1, \ldots, \mathcal{X}_r$. Then, with the notation of section 1.3,

$$\phi_\varepsilon^{(\mathcal{X}_1,\ldots,\mathcal{X}_r)*}[h+\varepsilon f] = h + \sum_{s=1}^{r} \varepsilon^s [h_{(s)} + f_{(s-1)}] + R(\varepsilon, r+1)[h+\varepsilon f]$$

So, to obtain H_ε in its normal form up to order r, say

$$\phi_\varepsilon^{(\mathcal{X}_1,\ldots,\mathcal{X}_r)*}[h+\varepsilon f] = h + \sum_{s=1}^{r} \varepsilon^s Z_s + R(\varepsilon, r+1)[h+\varepsilon f]$$

we have to solve, for each s=1,...,r, the equations

(1) $\quad h + f_{(s-1)} = Z_s$.

Since $h_{(s)} = \sum_{j=1}^{s} \frac{j}{s} :L_j h_{(s-j)}:$, $f_{(s-1)} = \sum_{j=1}^{s-1} \frac{j}{s-1} :L_j f_{(s-j-1)}:$, equations (1) can be rewritten, by isolating the addend $L_s h$ in $h_{(s)}$,

(2) $\quad \breve{L}_h \mathcal{X}_s = g_s - Z_s$

with, for each s=1,...,r,

$$g_s = f_{(s-1)} + [h_{(s)} - L_s h] = \sum_{j=1}^{s-1} :L_j [\frac{j}{s} h_{(s-j)} + \frac{j}{s-1} f_{(s-1-j)}]$$

In particular, $g_1 = f$.

Note that, for each s, g_s depends only on $h, f, \mathcal{X}_1, \ldots, \mathcal{X}_{s-1}$ (and their Poisson brackets) so that equations (2) can be solved recursively. When $h = \omega \cdot p$ with ω diofantine, each equation (2) has the solution

$$Z_s = \tilde{g}_{s,0}$$

$$\mathcal{X}_s(p,q) = \sum_{k \neq 0} \frac{\hat{g}_{s,k}(p)}{i k \cdot \omega} e^{i k \cdot q}$$

as stated in the following proposition, whose proof is deferred to Appendix A. Note that this proposition implies that the normalization procedure can be performed up to any finite order.

Proposition 2.1 Assume that $\omega \in R^n$ satisfies the diofantine condition $|\omega \cdot k| \geq \gamma |k|^{-n}$ for all $k \in \mathbb{Z}^n$, $k \neq 0$. Let f be an analytic function on a set $D = D_{\varrho,\sigma}(B)$, with $\varrho > 0$, $\sigma > 0$, B being an open subset of R^n. Consider positive numbers δ, ξ such that $\delta < \min(\varrho, 1)$, $\xi < \min(\sigma, 1)$, any integer $r \geq 1$ and define

(3)
$$\varepsilon_r^* = \frac{\varepsilon_1^*}{r^{2(n+1)}}$$
$$\varepsilon_1^* = \frac{\delta^2 \xi^{2(n+1)} \gamma}{2^{12+5n} n^n \|f\|_D}$$

Then for all $|\varepsilon| < \varepsilon_r^*$, the Hamiltonian $H_\varepsilon(p,q) = \omega \cdot p + \varepsilon f(p,q)$ is coniugated in $D-2(\delta,\xi)$ to its normal form

$$H'_\varepsilon(p,q) = \omega \cdot p + \sum_{s=1}^{r} \varepsilon^s Z_s(p) + R(r+1; H_\varepsilon)$$

via an analytic canonical transformation ϕ_ε which is real for real values of the variables and is $\frac{\varepsilon}{4\varepsilon_r^*}$ (δ,ξ)-close to the identity. The remainder $R(r+1;H_\varepsilon)$ is bounded, for each $|\varepsilon| < \varepsilon_r^*$, by

(4) $$\|R(r+1;H_\varepsilon)\|_{D-2(\delta,\xi)} \leq \left[\|h\|_D + \varepsilon_r^* \|f\|_D\right] \left(\frac{|\varepsilon|}{\varepsilon_r^*}\right)^{r+1}$$

or also (since $\varepsilon_r^* \|f\|_D \leq \gamma$):

(5) $\quad \| R(r+1;H_\varepsilon)\|_{D-\varrho(\delta,\xi)} \le [\|h\|_D + \gamma] \left(\frac{|\varepsilon|}{\varepsilon_r^*} \right)^{r+1}$

We use now this result to obtain an exponential estimate for the stability time (let us stress that from now on only real values of ε are of interest; to simplify the notation we will even think it to be positive). As an intermadiate, although not strictly necessary, step let us show, following [22], that exponentially small estimates for $\|R(r+1;H_\varepsilon)\|$ are obtained, for each ε, by choosing r so to minimize it. Let us look, for simplicity, only for the minimum of the factor

(6) $\quad G_\varepsilon(r) = \left(\frac{\varepsilon}{\varepsilon_r^*}\right)^r = \left(\frac{\varepsilon r^b}{\varepsilon_1^*}\right)^r$

(with $b = 2n+2$) of the r.h.s. of (5). Regarding r as a real variable it is easy to verify that G_ε reaches its minimum at

(7) $\quad \bar{r} = \bar{r}(\varepsilon) = \left(\frac{\varepsilon_0}{\varepsilon}\right)^{1/b}$

where

(8) $\quad \varepsilon_0 = e^{-b}\, \varepsilon_1^*$

and that then

(9) $\quad G_\varepsilon(\bar{r}) = e^{-b\bar{r}} = e^{-b(\varepsilon_0/\varepsilon)^{1/b}}$

To simplify the treatment, we impose that $\bar{r}(\varepsilon) \ge 1$, i.e. we restrict to $|\varepsilon| \le \varepsilon_0$. Define now, for each ε, the "optimal" normalization order $r_o(\varepsilon)$ as the integer part $[\bar{r}(\varepsilon)]$ of $\bar{r}(\varepsilon)$. Then $r_o(\varepsilon) \ge 1$ and so $\varepsilon^*_{r_o(\varepsilon)} = \varepsilon_1^* / r_o{}^b \le \varepsilon_1^*$; on the other hand, it is $\varepsilon^*_{r_0(\varepsilon)} \geq \varepsilon^*_1 / \bar{r}^{\,b} = e^b |\varepsilon|$ so that

$$(10) \qquad \frac{\varepsilon}{\varepsilon^*_{r_0(\varepsilon)}} \leq e^{-b}$$

which shows that the consistency condition $|\varepsilon| < \varepsilon^*_{r_0(\varepsilon)}$ is satified for each ε.

It only remains to estimate $G_\varepsilon(r_0) = (\varepsilon / \varepsilon^*_{r_0(\varepsilon)})^{r_0(\varepsilon)}$ or, more appropriately, the quantity $(\varepsilon / \varepsilon^*_{r_0})^{r_0+1}$ which appears in (5). Using (10), $r_0 + 1 \geq \bar{r}$ and $b > 1$ we have

$$\left(\frac{\varepsilon}{\varepsilon^*_{r_0(\varepsilon)}} \right)^{r_0+1} \leq e^{-b(r_0+1)} \leq e^{-b\bar{r}} \leq e^{-\bar{r}} = e^{-\left(\frac{\varepsilon_0}{\varepsilon}\right)^{1/b}}$$

It then follows that, in the same hypotheses of proposition 2.1, for each $|\varepsilon| \leq \varepsilon_0$ the normalization procedure can be performed until the the optimal order $r_0(\varepsilon)$ and then, from (5), one obtains

$$(11) \qquad \| R(r_0(\varepsilon)+1; H_\varepsilon) \|_{\delta - 2(\delta,\xi)} \leq \left[\|h\|_\delta + \gamma \right] e^{-\left(\frac{\varepsilon_0}{\varepsilon}\right)^{\frac{1}{2(n+2)}}}$$

This result is however insufficient to get an exponential estimate for the stability time. Indeed, while it gives such a bound for the image of the actions under the canonical transformation which normalizes H_ε, at the order $r_0(\varepsilon)$ the ε-closeness to the identity of this transformation is lost (see (10)).

It is anyway possible to obtain the sought estimate by getting somehow worse the bound (11) on the remainder $R(r+1;H_\varepsilon)$. Precisely, let us arrest the normalization procedure, for each $|\varepsilon| \leq \varepsilon_0$, at the order $r_m(\varepsilon) = [\bar{r}(\varepsilon)^{1/2}] \geq 1$. Since $\bar{r}(\varepsilon)^{1/2} = (\varepsilon_0/\varepsilon)^{1/2}$, we have $\varepsilon^*_{r_m} \geq \varepsilon^*_1/\bar{r}^{b/2} = e^b\sqrt{\varepsilon\varepsilon_0}$ and then

$$(12) \qquad \frac{\varepsilon}{\varepsilon^*_{r_m(\varepsilon)}} \leq e^{-b}\sqrt{\frac{\varepsilon}{\varepsilon_0}} < \sqrt{\frac{\varepsilon}{\varepsilon_0}} \quad ;$$

thus, an $\varepsilon^{1/2}$-closeness to the identity of the canonical transformation is now achieved. Moreover, it is now

$$\left(\frac{\varepsilon}{\varepsilon^*_{r_m(\varepsilon)}}\right)^{r_m+1} \leq \left(e^{-b}\sqrt{\frac{\varepsilon}{\varepsilon_0}}\right)^{\left(\frac{\varepsilon_0}{\varepsilon}\right)^{1/2b}} \leq \sqrt{\frac{\varepsilon}{\varepsilon_0}}\; e^{-\left(\frac{\varepsilon_0}{\varepsilon}\right)^{\frac{1}{2b}}}$$

where $b>1$ and $|\varepsilon| < \varepsilon_0$ have been used, so that we get from (5)

$$(13) \qquad \|R(r_m+1;H_\varepsilon)\|_{\Delta-2(\delta,\xi)} \leq \sqrt{\frac{\varepsilon}{\varepsilon_0}}\;\left[\|h\|_\Delta + \gamma\right]\; e^{-\left(\frac{\varepsilon_0}{\varepsilon}\right)^{\frac{1}{4(n+1)}}}$$

We can now state the final result:

Proposition 2.2 Assume that $\omega \in R^n$ satisfies $|\omega\cdot k| \geq \gamma|k|^{-n}$ for all $k \in \mathbb{Z}^n$, $k \neq 0$ and some $\gamma > 0$. Assume that f is a real analytic function on $B\times T^n$, $B \subset R^n$ being any open set, and that it has an analytic extension to $D_{\varrho,\sigma}(B)$, for some $0 < \varrho < 4$, $0 < \sigma < 4$. Consider the (real) Hamiltonian $H_\varepsilon = h + \varepsilon f$, with $h = \omega\cdot p$ and $|\varepsilon| < \varepsilon_n$, where

$$\varepsilon_n = \frac{\gamma\; \varrho^2\; \xi^{2(n+1)}}{2^{24+10n}\; e^{2(n+1)}\; n^n\; \|f\|_\Delta}$$

Then, for any initial condition $(p(0),q(0)) \in (B-\varrho/2)\times T^n$, the motion $(p(t),q(t))$ of the system described by H_ε satisfies

$$\| p(t) - p(0) \| \leq \frac{1}{8} \rho \sqrt{\frac{\varepsilon}{\varepsilon_m}}$$

for all

$$|t| \leq T = \frac{\rho \sigma}{2^5 [\|h\|_D + \gamma]} e^{\left(\frac{\varepsilon_0}{\varepsilon}\right)^{\frac{1}{4(n+1)}}}$$

Proof Let us use all the previous results and notations. Take $\delta = \rho/8$, $\xi = \sigma/8$ and note that then it is $\varepsilon_m = \varepsilon_0$, with ε_0 defined by (8). Take $|\varepsilon| < \varepsilon_m$ and apply proposition 2.1 with $r=r_m(\varepsilon)$; it then follows from (13) that, for all $|t| \leq T$ it is (with $b=2n+2$):

$$|t| \left\| \frac{\partial R(r_m+1; H_\varepsilon)}{\partial q} \right\|_{D-(2\delta,4\xi)} \leq$$

$$\leq |t| \frac{1}{2\xi} [\|h\|_D + \gamma] \sqrt{\frac{\varepsilon}{\varepsilon_m}} \quad e^{-(\varepsilon_0/\varepsilon)^{1/2b}} \leq \frac{1}{2} \sqrt{\frac{\varepsilon}{\varepsilon_m}} \delta$$

Toghether with the reality of the normalized Hamiltonian $H_\varepsilon \circ \phi_\varepsilon$ (for real values of the variables) this implies that, for each initial condition $\phi_\varepsilon(p(0),q(0)) \in [B-15\delta/4] \times T^n$ one has

$$\| \phi_\varepsilon p(t) - \phi_\varepsilon p(0) \| \leq \frac{1}{2} \delta \sqrt{\frac{\varepsilon}{\varepsilon_m}}$$

for all $|t| \leq T$. Since, by proposition 2.1 and (12), ϕ_ε is $\frac{1}{4} \sqrt{\frac{\varepsilon}{\varepsilon_m}} (\delta,\xi)$-closed to the identity we have, for $|t| \leq T$ and $(p(0),q(0)) \in (B-4\delta) \times T^n$:

$$\|p(t)-p(0)\| \le \|p(t)-\phi_\varepsilon p(t)\| + \|\phi_\varepsilon p(t)-\phi_\varepsilon p(0)\| + \|\phi_\varepsilon p(0)-p(0)\|$$

$$\le \frac{\delta}{4}\sqrt{\frac{\varepsilon}{\varepsilon_\mu}} + \frac{\delta}{2}\sqrt{\frac{\varepsilon}{\varepsilon_\mu}} + \frac{\delta}{4}\sqrt{\frac{\varepsilon}{\varepsilon_\mu}} = \delta\sqrt{\frac{\varepsilon}{\varepsilon_\mu}} .$$

2.4 NORMALIZATION WITH CUTOFF

We present in this last section the normalization procedure, employing a cutoff, for a nearly integrable Hamiltonian in a complex neighbourhood of one of the blocks B_κ which, as discussed in 2.2.D, constitutes the "analytic" part of the proof of Nekhoroshev's theorem. After a brief discussion of the normalization procedure at a formal level we state, in proposition 2.3, its very possibility; the proof is deferred to appendix B. To give a concrete example of this procedure we apply this result to the case, considered in the previous section, of a system of harmonic oscillators with a diofantine frequency; by minimizing the remainder, at each fixed ε, with respect to both the normalization order and the cutoff, an exponential estimate for it is again obtained (although sligtly worse than that obtained without the use of the cutoff). We do not state here a stability result similar to proposition 2.2, which could be obtained and proved just in the same way.

Consider an Hamiltonian $H_\varepsilon(p,q) = h(p)+\varepsilon f(p,q)$, analytic on a set $D = D_{\rho,\sigma}(B)$, and decompose it as $H_\varepsilon = H_\varepsilon^{\le K} + H_\varepsilon^{>K}$, for some positive integer K. Employing a product of r Lie series, we want to bring $H_\varepsilon^{\le K} = h+\varepsilon f^{\le K}$ into its normal form, up to order r, adapted to a lattice $\mathcal{K}$, say

$$\phi_\varepsilon^{(\mathcal{X}_1,\dots,\mathcal{X}_r)*}[h+\varepsilon f^{\leq K}] = h + \sum_{s=1}^{r} \varepsilon^s Z_s + R(r+1,\varepsilon)[h+\varepsilon f^{\leq K}]$$

with $Z_s = \Pi_{\mathcal{K}} Z_s$. To this purpose, proceeding exactly as in the previous section, we have to solve, for each s=1,..,r, the equation

$$(1) \qquad L_h \mathcal{X}_s = g_s - Z_s$$

where now

$$(2) \qquad g_s = (f^{\leq K})_{(s-1)} + [h_{(s)} - L_s h].$$

The essential point is that the property of the perturbation $f^{\leq K}$ of having finitely many nonzero Fourier components is preserved at any order, so that -if a suitable nonresonance condition is satisfied- each equation (2) has an analytic solution

$$(3) \qquad \begin{aligned} Z_s &= \Pi_{\mathcal{K}}\, g_s \\ \mathcal{X}_s(p,q) &= \sum_{k\notin\mathcal{K}} \frac{\tilde{g}_{s,k}(p)}{i k\cdot\omega(p)}\, e^{ik\cdot q} \end{aligned}$$

Indeed, it is easy to see that, if F and F' are two functions such that $\tilde{F}_k = 0$ for all $|k|>K$ and $\tilde{F}'_k = 0$ for all $|k|>K'$, then $\{F,F'\}$ has nonzero Fourier components only for $|k|\leq K+K'$. A simple induction (see [8]) then shows that, being $\tilde{h}_s = 0$ for $|k|>0$ and $\tilde{g}_{1,k} = \widetilde{(f^{\leq K})}_k = 0$ for $|k|>K$, the functions g_s, Z_s, $\mathcal{X}_s$ (formally) defined by (2) and (3) have the following properties: for each $s\geq 1$, $\mathcal{X}_s$ has nonzero Fourier components only for

$|k| \le (s-1)K$ while g_s, and thus Z_s, only for $|k| \le sK$.

Thus, in order to solve the r equations (1) with (3), we have to require that $\omega = \frac{\partial h}{\partial p}$ satisfies, at each point $p \in D_\varrho(B)$, a non-resonance condition

$$|\omega(p)\cdot k| \ge \alpha(r,K) \quad \text{for all } k \notin \mathcal{M},\ |k| \le rK$$

for some positive $\alpha(r,K)$. By taking into account also the "ultraviolet" part of the perturbation $f^{>K}$, in the way outlined in 2.2.D, and performing all the necessary estimates, we arrive at the following

Proposition 2.3 Let $B \subset R^n$ be an open set and ϱ, σ positive numbers. Assume that h is analytic on $D = D_{\varrho,\sigma}(B)$ and there exist a lattice $\mathcal{M} \subset \mathbb{Z}^n$ of dimension $0 \le m < n$, positive integers r and K and a positive $\alpha = \alpha(r,K)$ such that, for all $p \in D_\varrho(B)$ one has

$$(4) \qquad |\omega(p)\cdot k| \ge \alpha(r,K) \qquad \text{for all } k \notin \mathcal{M},\ |k| \le rK$$

Consider a function f analytic on D, positive numbers $\delta < \min(\varrho,1)$, $\xi < \min(\sigma,1)$ and define

$$\varepsilon^*_{r,K} = \frac{\delta^2\, \xi^{2(n+1)}}{2^{8n+11}\ \|f\|_D} \cdot \frac{\alpha(r,K)}{r^{n+2}}$$

Then, for all $|\varepsilon| < \varepsilon^*_{r,K}$, there exists an analytic canonical diffeomorphism ϕ_ε of $D-2(\delta,\xi)$ onto its image which is $\frac{\varepsilon}{4\varepsilon^*_r}$

(δ, ξ)-close to the identity and gives $H_\varepsilon = h + \varepsilon f$ the form

$$\phi_\varepsilon^* [h + \varepsilon f] = h + \sum_{s=1}^{r} \varepsilon^s Z_s + R(r,K;h+\varepsilon f)$$

with $\Pi_\mathcal{R} Z_s = Z_s$ and

$$(5) \qquad \| R(r,K;h+\varepsilon f) \|_{D-2(\delta,\xi)} \le \frac{\varepsilon}{\varepsilon^*_{r,K}} \left[\left(\|h\|_D + \varepsilon^*_{r,K} \left(\frac{4}{\xi}\right)^n \|f\|_D \right] \left(\frac{\varepsilon}{\varepsilon^*_{r,K}} \right)^r + \right.$$

$$\left. + \; \varepsilon^*_{r,K} \left(\frac{8}{\xi}\right)^n \|f\|_D \, e^{-K\xi/2} \right]$$

The proof of this proposition is deferred to appendix B.

We now apply this result to the case $h = \omega \cdot p$, with ω satisfying the diofantine condition $|\omega \cdot k| \ge \gamma |k|^{-n}$ for all $k \ne 0$. In this case, (4) is obviously satisfied with

$$\alpha(r,K) = \gamma \, (rK)^{-n}$$

and we can rewrite

$$\varepsilon^*_{r,K} = \frac{\varepsilon^*_{1,1}}{r^{2n+2} \, K^n}$$

$$\varepsilon^*_{1,1} = \frac{\delta^2 \, \xi^{2(n+1)} \, \gamma}{2^{8n+14} \, \|f\|_D}$$

We want now to choose r and K, for a fixed ε, in order to minimize the remainder $\|R(r,K;h+\varepsilon f)\|$. To simplify the computations, let us assume that $\gamma \le 1$ and note that for $r \ge 1$ and $K \ge 1$ it is $\varepsilon^*_{r,K} \, (8/\xi)^n \, \|f\|_D \le \gamma/4 \le \gamma n/(2n+2) \le \gamma$, so that (5) implies

$$(6) \qquad \|R(r,K;H_\varepsilon)\|_{D-2(\delta,\xi)} \le \frac{\varepsilon}{\varepsilon^*_{r,K}} \left[\|h\|_D + \gamma\right] \left[\left(\frac{\varepsilon}{\varepsilon^*_{r,K}}\right)^r + \frac{n}{2n+2} e^{-\frac{K\xi}{2}} \right]$$

Let us now write $b=2n+2$, $c=\xi/2$ and

$$U_\varepsilon(r,K) = G_\varepsilon(r,K) + \frac{n}{b} e^{-Kc}$$

$$G_\varepsilon(r,K) = \left(\frac{\varepsilon r^b K^n}{\varepsilon^*_{1,1}} \right)^r$$

The extremum conditions $\frac{\partial U_\varepsilon}{\partial r}(\bar{r},\bar{K}) = \frac{\partial U_\varepsilon}{\partial K}(\bar{r},\bar{K}) = 0$ give

$$\bar{r}\,\bar{K}^{n/b} = \frac{1}{e} \left(\frac{\varepsilon^*_{1,1}}{\varepsilon} \right)^{1/b}$$

$$G_\varepsilon(\bar{r},\bar{K}) \cdot \frac{n\bar{r}}{\bar{K}} = \frac{nc}{b} e^{-c\bar{K}}$$

The first one implies that $G_\varepsilon(\bar{r},\bar{K}) = e^{-b\bar{r}}$ and then the second one becomes $b\cdot\bar{r}\cdot e^{-b\bar{r}} = c\cdot\bar{K}\cdot e^{-c\bar{K}}$, i.e. (being $r\ge1$)

$$(7) \qquad b\cdot\bar{r} = c\cdot\bar{K}$$

It then follows that

$$\bar{r} = \left(\frac{\varepsilon_0}{\varepsilon}\right)^{\frac{1}{3n+2}}$$

$$\bar{K} = \frac{2n+2}{n}\,\bar{r}$$

where

$$\varepsilon_0 = \frac{\delta^2 \xi^{3n+2} \gamma}{e^{2n+2} 2^{10n+14} (n+1)^n}$$

For all $|\varepsilon| < \varepsilon_0$ it is $\bar{r}(\varepsilon) \geq 1$ and then also $\bar{K}(\varepsilon) \geq 1$. Moreover, U_ε has a minimum at $(\bar{r}(\varepsilon), \bar{K}(\varepsilon))$. Let us take $r_0 = [\bar{r}]$, $K_0 = [\bar{K}]$. Proceeding as in the previous section (and using (4)) we get

$$U_\varepsilon(r_0, K_0) \leq e^{-b(1+r_0)} + \frac{n}{b} e^{c-b} e^{c-c(K_0+1)} \leq$$

$$\leq e^{-b\bar{r}} + \frac{n}{b} e^{c-b} e^{-c\bar{K}} = e^{b\bar{r}} \left[1 + \frac{n}{b} e^{c-b}\right]$$

and then, using $b > c$ and $n < b$,

$$U_\varepsilon(r_0, K_0) \leq 2e^{-b\bar{r}} \leq 2e^{-\bar{r}}$$

Thus, we finally get

$$\| R(r_0, K_0; H_\varepsilon) \|_{D-2(\delta,\xi)} \leq 2\left[\|h\|_D + \gamma\right] e^{-\left(\frac{\varepsilon_0}{\varepsilon}\right)^{\frac{1}{3n+2}}}$$

APPENDIX A

PROOF OF PROPOSITION 2.1

A.1 NOTATION

In this appendix we give a complete proof of the proposition 2.1 of section 2.3, which states the existence of the normal form up to any finite order r for the Hamiltonian of a system of weakly coupled and strongly nonresonant harmonic oscillators. The general lines of the present proof are those of [22], but a number of differences comes from the use of a different method to generate the canonical transformation.

The strategy is the following:

(a) we use a product of Lie series to normalize the Hamiltonian, applying then the formal scheme of section 2.3 which gives, at each "step" s=1,...,r, the equation $L_h \chi_s = g_s - \widetilde{g}_{s,0}$;

(b) it is a classic result that, if g is analytic then this equation has an analytic solution (proposition A.1);

(c) proceeding recursively, we show that all the $g_1,\dots,g_r$, and then $\chi_1,\dots,\chi_n$, are analytic and that,

(d) moreover, the χ_s satisfy an estimate of the kind $\|\chi_s\| \le A^s$ (proposition A.3; this is the central part of the proof);

(e) the use of proposition 1.2 of section 1.4 then immediately gives the existence of the canonical transformation.

For convenience, we collect here the notations and the hypotheses to be used throughout this Appendix. We write $D = D_{\rho,\sigma}(B)$, B being an open set of R^n and ρ,σ some positive numbers; δ and ξ are positive numbers such that $\delta < \min(\rho,1)$ and $\xi < \min(\sigma,1)$; r is some positive integer. We will use the following subsets of D: for all $j=1,\ldots,r+1$:

$$D_j = D - \frac{j-1}{r}(\delta,\xi)$$

(note that $D_1 = D$) and for all $j=1,\ldots,r$:

$$D'_j = D_j - \frac{1}{2r}(\delta,\xi).$$

The following condition will be referred to as (HA1): $h = \omega\cdot p$ with $\omega \in R^n$ is such that $|\omega\cdot k| \geq \gamma |k|^{-n}$ for all $k \in \mathbb{Z}^n$, $k \neq 0$, γ being some positive number.

A.2 SOLUTION OF THE PERTURBATIVE EQUATION

We first show the existence of an analytic solution χ_s of the equation $L_h \chi_s = g_s - \tilde{g}_{s,0}$ when g_s is analytic. This is a well known result.

Proposition A.1 Let h be as in (HA1). Assume that g_s is analytic on D_s, for any $1 \leq s \leq r$. Then the equation $L_h \chi_s = g_s - \tilde{g}_{s,0}$ has a solution χ_s which is analytic on D'_s and satisfies

(1) $\quad \|\chi_s\|_{D'_s} \leq M \|g_s\|_{D_s}$

with

(2) $$M = \frac{1}{\gamma}\left(\frac{2^5 n r^2}{\xi^2}\right)^n$$

Furthermore

(3) $$\| L_h \mathcal{X}_s \|_{D'_s} \leq 2 \| g_s \|_{D_s}$$

Proof Define $\mathcal{X}_s(p,q) = \sum_{k\neq 0} \frac{\tilde{g}_{s,k}}{i k\cdot\omega} e^{i k\cdot q}$ so that

$$\| \mathcal{X}_s \|_{D'_s} \leq \sum_{k\neq 0} \frac{\| \tilde{g}_{s,k} \|_{D_s}}{|\omega\cdot k|} e^{|k|\left[\sigma - \frac{s-1}{r}\xi - \frac{1}{2r}\xi\right]}$$

Using lemma 2.1 to estimate $\| \hat{g}_{s,k} \|_{D_s}$ and the diofantine condition (HA1) this gives

$$\| \mathcal{X}_s \|_{D'_s} \leq \frac{1}{\gamma} \| g_s \|_{D_s} \sum_{k\neq 0} |k|^n e^{-|k|\frac{\xi}{2r}}$$

Note now that, for all $a,b,c > 0$ it is

$$a^b \leq \left(\frac{b}{ec}\right)^b e^{ac} \qquad ;$$

indeed, this inequality is equivalent to $ac/b \leq \exp(ab/c \ -1)$ which is evidently true since $x \leq \exp(x-1)$ for all $x \in R$. So, with $a=|k|$, $b=n$, $c=\xi/4r$, we get $|k|^n \leq (4nr/e\xi) \exp(|k|\xi/4r)$. It follows that

$$\| \mathcal{X}_s \|_{D'_s} \leq \frac{1}{\gamma} \| g_s \|_{D_s} \left[\frac{4nr}{e\xi}\right]^n \sum_{k\neq 0} e^{-|k|\frac{\xi}{4r}}$$

Evaluating the last sum like in the proof of lemma 2.1 we get

$$\| \mathcal{X}_s \|_{D'_s} \leq \| g_s \|_{D_s} \frac{1}{\gamma}\left[\frac{2^6 n r^2}{e\xi^2}\right]^n \leq M \| g_s \|_{D_s}$$

The analyticity of $\mathcal{X}_s$ on D_s' follows from this if g_s is bounded on D_s; otherwise, it is proven by repeating the same procedure on some subset of D_s. Finally, (3) follows from $\|L_h \mathcal{X}_s\|_{D'_s} \leq \|g_s - \tilde{g}_{s,0}\|_{D_s} \leq \|g_s\|_{D_s} + \|\hat{g}_{s,0}\|_{D_s}$ and $\|\tilde{g}_{s,0}\|_{D_s} \leq \|g_s\|_{D_s}$.

A.3 RECURSIVE ESTIMATE OF THE $f_{(s)}$

We now give a recursive estimate for the first r+1 terms $f_{(0)},\dots,f_{(r)}$ of the transformed of a function f under a product of r Lie series, which will be used in the following section. Because of the ordering, it is not possible to directly estimate $f_{(s)} = \sum_{j=1}^{s} \frac{j}{s} :L_j f_{(s-j)}:$ as $\|f_{(s)}\| \leq \sum_{j=1}^{s} \frac{j}{s}\beta \|f_{(s-j)}\| \|\mathcal{X}_j\|$ with some constant β; however, it is possible to give a similar relation for suitable majorants of the $\|f_{(s)}\|$.

Proposition A.2 Consider any integer $s \geq 1$ and, for each $j=1,\dots,s$, a function $\mathcal{X}_j$ analytic on D_j'. Define recursively the numbers

$$S_0 = 1$$

(4)

$$S_p = \frac{16 r^2}{\delta \xi} \sum_{j=1}^{p} \|\mathcal{X}_j\|_{D'_j} S_{p-j} \qquad (p=1,\dots,s)$$

where r is any integer $\geq s$. Let f be a function analytic on D and consider the product of Lie series for f generated by $\mathcal{X}_1,\dots,\mathcal{X}_r$, written $\sum_{p\geq 0} \varepsilon^p f_{(p)}$. Then, for all $0 \leq p \leq s$, $f_{(p)}$ is analytic on D_{p+1} and satisfies

(5) $$\|f_{(p)}\|_{D_{p+1}} \leq S_p \|f\|_D$$

Proof The analyticity of $f_{(p)}$ immediately follows from the hypotheses. For p=0, (5) is true by definition. Let $1 \leq p \leq s$. By lemma 1.3 (with δ, ξ replaced by $p\delta/r$, $p\xi/r$) it follows that, for all $\nu \in \mathbb{N}_0^p$, $\nu \neq 0$, it is

$$\left\| \frac{L_p^{\nu_p} \cdots L_1^{\nu_1}}{\nu_p! \cdots \nu_1!} f \right\|_{D_{p+1}} \leq \|f\|_D \prod_{j=1}^{p} \left(\frac{16\, r^2}{\delta \xi}\right)^{\nu_j} \|X_j\|_{D'_j}^{\nu_j}.$$

Let us define $S(\nu) = \prod_{j=1}^{p} \left(\frac{16\, r^2}{\delta \xi}\right)^{\nu_j} \|X_j\|_{D'_j}^{\nu_j}$. Then

$$\|f_{(p)}\|_{D_{p+1}} \leq \|f\|_D \sum_{\nu \in P(p,s)} S(\nu)$$

So, to prove (5) it is sufficient to show that, for all $1 \leq p \leq s$ one has

(6) $$S_p \geq \sum_{\nu \in P(p,s)} S(\nu)$$

It is easy to verify that this is true for p=1. Proceeding by induction, we assume that it is also true for all $p=2,\ldots,q-1$, any $q \leq s$, and show that then it is satisfied for $p=q$, too. By the definition (4) and the induction hypothesis it follows that

$$S_q \geq \sum_{j=1}^{q} \sum_{\nu \in P(q-j;s)} \frac{16\, r^2}{\delta \xi} \|X_j\|_{D'_j} S(\nu)$$

For each $j \geq 1$, we denote $e_j = (\delta_{1j}, \ldots, \delta_{sj})$ the j-th unit vector of R^s. It follows from the definition of $S(\nu)$ that $S(\nu+e_j) = (4r^2/\delta\xi)\, \|\mathcal{X}_j\|_{D'_j}\, S(\nu)$, so that

$$S_q \geqslant \sum_{j=1}^{q} \sum_{\nu \in P(q-j;s)} \overline{} \quad S(\nu+e_j).$$

We now perform, for each $j=1,\ldots,q$, the change of summation index $\nu \mapsto \nu + e_j = \nu'$. Since $\nu \in P(q-j;s)$ iff $\nu' = \nu+e_j \in P(q;s)$ and $\nu'_j \geq 1$ we get

$$S_q \geqslant \sum_{j=1}^{q} \sum_{\substack{\nu \in P(q;s) \\ \nu_j \geqslant 1}} \overline{} \quad S(\ ).$$

But since each $\nu \in P(q;s)$ has at least one of the first q components different from zero, the r.h.s. of the last inequality is not lesser than $\sum_{\nu \in P(q;s)} \overline{}\, S(\nu)$ so that (6), and so (5), is proved. $\blacksquare$

A.4. ESTIMATES OF THE g_s AND OF THE $\mathcal{X}_s$

We now complete the estimate of the generating functions by obtaining an estimate for each $\|g_s\|_{D_s}$. From this and proposition A.1 then follows an estimate for the $\mathcal{X}_s$. The final result is the following.

Proposition A.3 Assume that h is as in (HA1), f is a function analytic on D and $r \geqslant 1$ is any integer. Then it is possible to define the functions $\mathcal{X}_1, \ldots, \mathcal{X}_r$ through the recursive scheme

$$\chi_s(p,q) = \sum_{k\neq 0} \frac{\tilde{g}_{s,k}(p)}{i\,k\cdot\omega}\, e^{i k\cdot q}$$

$$g_1 = f$$

$$g_s = f_{(s-1)} + [h_{(s)} - L_s\, h] \qquad (s=1,\dots,r)$$

where $L_j = L_{\chi_j}$, $f_{(0)}=f$, $f_{(s)} = \sum_{\nu\in P(s)} \frac{L_s^{\nu_s}\cdots L_1^{\nu_1}}{\nu_s!\cdots\nu_1!} f$ and the $h_{(s)}$ are similarly defined. Moreover,

(a) each g_s is analytic on D_s and satisfies

(7) $$\|g_s\|_{D_s} \leq A^{s-1}\, \|f\|_D$$

where (with M defined by (2))

(8) $$A = \frac{2^7 r^2}{\delta\xi}\, M\, \|f\|_D\ ;$$

(b) each χ_s is analytic on D'_s and satisfies

(9) $$\|\chi_s\|_{D'_s} \leq A^s \quad .$$

To the proof of this proposition is devoted all this section. We articulate it in a number of steps. Of course, only (a) has to be proved, since (b) follows from it and proposition A.1.

Let us recall that, by definition, $g_s = f_{(s-1)} + [h_{(s)} - L_s h]$. So, for all $s \geqslant 2$ it is

(10) $$g_s = f_{(s-1)} + \sum_{\substack{\nu\in P(s)\\ \nu_s=0}} \frac{L_s^{\nu_s}\cdots L_1^{\nu_1}}{\nu_s!\cdots\nu_1!}\, h$$

This is also true for s=1, with the convention (always understood) that a sum over an empty set of indices is equal to zero. We have to use (10) to recursively estimate the g_s. To this purpose, it is convenient to rewrite (10) by collecting toghether in the last sum all those ordered monomials which terminate on the right with the same L_j, namely correspon-ding to some $\vartheta \in P(s)$ having the first $j-1$ components zero and the j-th one different from zero.

Let us introduce some notation to be used to this purpose. For each $s \geq 1$ and $1 \leq j \leq s$, we denote $P_j(s)$ the subset of all these ϑ, i.e.

$$P_j(s) = \{ \vartheta = (\vartheta_1, \dots, \vartheta_s) \in P(s): \vartheta_j \geq 1 \text{ and, if } j \geq 2, \vartheta_p = 0 \text{ for } 1 \leq p \leq j \} . \tag{11}$$

Note that $P_1(1) = \phi$. Moreover, for each $s \geq 1$ it is

$$\bigcup_{1 \leq j \leq [s/2]} P_j(s) = \{ \vartheta \in P(s) : \vartheta_s = 0 \} \tag{12}$$

where [ ] denotes the integer part. For $s=1$ this is true since both the sets are empty; for $s \geq 2$, it follows from $\sum_{j=1}^{s} j \vartheta_j = s$ that if $\vartheta \in P(s)$ and $\vartheta_s = 0$ then there exists at least one $j \leq [s/2]$ such that $\vartheta_j \geq 1$ (in other words, $P_j(s) = \phi$ for $j > [s/2]$). Finally, note that, denoting (for each $s \geq 1$) by $\hat{\vartheta}^{(s)}$ the unique vector of $P(s)$ having only the last component different from zero, i.e. $\hat{\vartheta}^{(s)} = (0, \dots, 0, 1)$, one has

$$P(s) = \left\{ \bigcup_{1 \leq j \leq [s/2]} P_j(s) \right\} \cup \left\{ \hat{\vartheta}^{(s)} \right\} \tag{13}$$

According to (11) and (12), (10) can be written in the form

$$(14) \qquad g_s = f_{(s-1)} + \sum_{j=1}^{[s/2]} \sum_{\vartheta \in P_j(s)} \frac{L_{s-j}^{\vartheta_{s-j}} \cdots\cdots L_j^{\vartheta_j - 1}}{\vartheta_{s-j}! \cdots\cdots \vartheta_j!} (L_j h)$$

which we will now use to estimate g_s in terms of $g_1,\dots,g_{s-1}$ and $f_{(s-1)}$. The crucial point is not to estimate the last $L_j h$ as a Poisson bracket, but instead to estimate it with (3) of proposition A.1. In this way we arrive at the following

Lemma A.1 Assume that $h, f, g_j, \mathcal{X}_j, f_{(j)}$ and $h_{(j)}$ are as in proposition A.3. Then, for all $s=1,\dots,r$, g_s is analytic on D_s and satisfies

$$(15) \qquad \|g_s\|_{D_s} \le \|f_{(s-1)}\|_{D_s} + \sum_{j=1}^{[s/2]} \sum_{\vartheta \in P_j(s)} \left(\frac{16\, r^2 M}{\delta \xi}\right)^{|\vartheta|-1} \prod_{p=j}^{s-j} \|g_p\|_{D_p}^{\vartheta_p}$$

where $|\vartheta| = \sum_j \vartheta_j$ and M is defined by (3). Furthermore, for all $1 \le s \le r$, $\mathcal{X}_s$ is analytic on D_s' and satisfies (1), while $f_{(s)}$ is analytic on D_{s+1} and satisfies (5).

Proof We proof the analyticity of g_s on D_s and (15) by induction on s. Clearly, they are both true for s=1, since $g_1 = f$. Let us assume they are also true for all $j=2,\dots,s-1$, any $s \ge 2$.

Under this hypothesis, $\mathcal{X}_j$ is analytic on D_j' and satisfies (2), for all $1 \le j \le s-1$, as stated in proposition A.1; moreover, by proposition A.2, also $f_{(s-1)}$ is analytic on D_s. Then g_s is analytic on D_s, being a finite sum of analytic functions.

To derive (15) it is sufficient to estimate the monomials

$L_{s-j}^{\vartheta_{s-j}} \dots L_j^{\vartheta_j - 1}(L_j h)$ for $\vartheta \in P_j(s)$. Proceeding in the way described in section 1.4, we perform the overall estimate on D_s, enlarging the domain of $(\delta/2r, \xi/2r)$ at each step; it is easy to recognize that in this way each χ_p is estimated on a domain contained in $D_p^{\cdot} - (\delta/2r, \xi/2r)$, so that we can controll it with $\|\chi_p\|_{D'_p}$. So, for each $\vartheta \in P_j(s)$ and $1 \le j \le [s/2]$:

$$\left\| \frac{L_{s-j}^{\vartheta_{s-j}} \cdots L_j^{\vartheta_j}}{\vartheta_{s-j}! \cdots \vartheta_j!} h \right\|_{D_s} \le \frac{1}{\vartheta_j} \left(\frac{16 r^2}{\delta \xi} \right)^{|\vartheta|} \|\chi_{s-j}\|_{D'_{s-j}}^{\vartheta_{s-j}} \cdots \|\chi_j\|_{D'_j}^{\vartheta_j - 1} \|L_j h\|_{D'_j}$$

From this, (15) is obtained by using inequalities (1) for $1 \le p < s$ and (3) for $1 \le j \le [s/2]$ and noting that $1/\vartheta_j \le 1$ for all $\vartheta \in P_j^{\cdot}(s)$. Finally, this implies via proposition A.1 that χ_s is analytic on $D_s^{\cdot}$ and satisfies (1); in turn, by proposition A.2, $f_{(s)}$ is analytic on D_{s+1} and satisfies (5). ∎

Note that this lemma implies the very feasibility of the normalization procedure up to any order r, provided the g_s do not explode. So, we have now to obtain explicit estimates for them by working out (15). We first reduce it to an inequality of the type $\|g_s\|_{D_s} \le \mathrm{cost} \cdot \mathcal{G}_s \cdot \|f\|_D$, where the $\mathcal{G}_s$ are numbers. Although the $\mathcal{G}_s$ are defined by an (apparently) cumbersome recursion, this is important since it singles out the correct dependence of the $\|g_s\|$ on the parameters (this procedure is basically the same used in [22]).

Lemma A.2 Assume the same hypotheses and use the same notations of the previous lemma. Define the numbers Ξ_j, $\mathcal{G}_j$, j=1,...,r, via the recursive formulas

$$(16)\quad \begin{cases} \mathcal{F}_1 = \mathcal{G}_1 = 1 \\ \mathcal{F}_s = \sum_{j=1}^{s-1} \mathcal{F}_j \, \mathcal{G}_{s-j} \\ \mathcal{G}_s = \mathcal{F}_s + \sum_{j=1}^{[s/2]} \sum_{\vartheta \in P_j(s)} \mathcal{G}_j^{\vartheta_j} \cdots \mathcal{G}_{s-j}^{\vartheta_{s-j}} \end{cases} \qquad (s \geq 2)$$

Then, for all s=1,...,r it is

$$(17)\qquad \| g_s \|_{D_s} \leq \mathcal{G}_s \, B^{s-1} \, \| f \|_D$$

with

$$(18)\qquad B = \frac{16 r^2}{\delta \xi} \, M \, \| f \|_D$$

where M is defined as in (2).

Proof According to the previous lemma, for all s=1,...,r, the $f_{(s)}$ satisfy (3), i.e.

$$(3)\qquad \| f_{(s-1)} \| \leq S_{s-1} \, \| f \|_D$$

with the S_j defined by (4), namely $S_0 = 1$ and

$$(4)\qquad S_{s-1} = \frac{16 r^2}{\delta \xi} \sum_{j=1}^{s-1} \| \chi_j \|_{D'_j} \, S_{s-1-j} \qquad (s=2,3,\ldots)$$

To prove (17) it is then enough to show that, if $\mathcal{F}_j$ and $\mathcal{G}_j$ are defined as in (16), then for all j=1,..,r it is

$$(19)\qquad S_{j-1} \leq \mathcal{F}_j \, B^{j-1}$$

$$(20)\qquad \| g_j \|_{D_j} \leq \mathcal{G}_j \, B^{j-1} \, \| f \|_D$$

For j=1 they hold by definition; they are also true for j=2, as one may verify using $\|\mathcal{S}_1\|_{D_1'} \le M\,\|g_1\|_{D_1} = M\,\|f\|_D$. Proceeding by induction, we assume that (19) and (20) hold for all $j \le s-1$, any $2 \le s \le r$. We first show that then (19) is true also for $j=s$; from (4), using (1), it follows that

$$S_{s-1} \le \frac{16\,r^2}{\delta\,\xi}\, M \sum_{j=1}^{s-1} \|g_j\|_{D_j}\, S_{s-1-j}$$

and so, for the induction hypothesis,

$$S_{s-1} \le \frac{16\,r^2}{\delta\,\xi}\, M\, \|f\|_D\, B^{s-2} \sum_{j=1}^{s-1} \mathcal{G}_j\, \mathcal{F}_{(s-j)}$$

which, by the very definition of B and $\mathcal{F}_s$, coincides with (19). We now show that also (20) is true for $j=s$. Inserting (3) into (15) we have

$$\|g_s\|_{D_s} \le S_{s-1}\,\|f\|_D + \sum_{j=1}^{[s/2]} \sum_{\nu \in P_j(s)} \left(\frac{16\,r^2}{\delta\,\xi}\, M\right)^{|\nu|-1} \prod_{p=j}^{s-j} \|g_p\|_{D_p}^{\nu_p}$$

and then, by the induction hypothesis,

$$\|g_s\|_{D_s} \le \mathcal{F}_s\, B^{s-1}\, \|f\|_D + \sum_{j=1}^{[s/2]} \sum_{\nu \in P_j(s)} \left(\frac{16\,r^2}{\delta\,\xi}\, M\right)^{|\nu|-1} \|f\|_D^{\nu} \cdot$$

$$\cdot\, B^{(s-j-1)\,\nu_{s-j} + \dots + (j-1)\,\nu_j}\, \mathcal{G}_{s-j}^{\nu_{s-j}} \cdots \mathcal{G}_j^{\nu_j}$$

Observing that $(s-j-1)\,\nu_{s-j} + \dots + (j-1)\,\nu_j = s - |\nu|$ for all $\nu \in P(s)$ and taking into account the definition of B we get

$$\|g_s\|_{D_s} \le \|f\|_D\, B^{s-1} \left\{ \mathcal{F}_{(s)} + \sum_{j=1}^{[s/2]} \sum_{\nu \in P_j(s)} \mathcal{G}_{s-j}^{\nu_{s-j}} \cdots \mathcal{G}_j^{\nu_j} \right\}$$

which, because of the definition (16), is (20).

Although the last inequality (16) looks very cumbersome, it can be greatly simplified:

Lemma A.3 The numbers $\mathcal{G}_1, \ldots, \mathcal{G}_r$ defined as in (16) satisfy

$$(21) \quad \begin{cases} \mathcal{G}_1 = 1 \\ \mathcal{G}_s \leq 2 \sum_{j=1}^{s-1} \mathcal{G}_j \mathcal{G}_{s-j} \end{cases}$$

Proof Take any s≥2. Note that, according to (12) and (13), the definition (16) of $\mathcal{G}_p$ (p≥1) can be rewritten in the form

$$\mathcal{G}_p = \mathcal{F}_p + \sum_{\substack{\vartheta \in \mathbb{P}(p) \\ \vartheta_p = 0}} \mathcal{G}_1^{\vartheta_1} \cdots\cdots \mathcal{G}_{p-1}^{\vartheta_{p-1}}$$

or also

$$(22) \quad \mathcal{G}_p = \mathcal{F}_p + \sum_{\vartheta \in \mathbb{P}(p)} \mathcal{G}_1^{\vartheta_1} \cdots\cdots \mathcal{G}_p^{\vartheta_p} - \mathcal{G}_p$$

Since this is obviously true also for p=1, it follows that

$$\sum_{j=1}^{s-1} \mathcal{G}_j \mathcal{G}_{s-j} = \sum_{j=1}^{s-1} \mathcal{G}_j \left\{ \mathcal{F}_{s-j} + \sum_{\vartheta \in \mathbb{P}(s-j)} \mathcal{G}_1^{\vartheta_1} \cdots \mathcal{G}_{s-j}^{\vartheta_{s-j}} - \mathcal{G}_{s-j} \right\}$$

and so, taking into account the definition of $\mathcal{F}_s$:

$$2 \sum_{j=1}^{s-1} \mathcal{G}_j \mathcal{G}_{s-j} = \mathcal{F}_s + \sum_{j=1}^{s-1} \sum_{\vartheta \in \mathbb{P}(s-j)} \mathcal{G}_1^{\vartheta_1} \cdots \mathcal{G}_j^{1+\vartheta_j} \cdots \mathcal{G}_{s-j}^{\vartheta_{s-j}}$$

For each $j=1,..,s-1$, perform now the change of summation index $\vartheta \mapsto \vartheta + e_j = \vartheta'_j$ (see the proof of proposition A.2); then

$$2\sum_{j=1}^{s-1} \mathcal{G}_j \mathcal{G}_{s-j} = \mathcal{F}_s + \sum_{j=1}^{s-1} \sum_{\substack{\vartheta\in P(s)\\ \vartheta_j \geq 1}} \mathcal{G}_1^{\vartheta_1} \cdots \mathcal{G}_s^{\vartheta_s}$$

Since all the $\vartheta \in P(s)$ but $\tilde{\vartheta}^{(s)} = (0,..,0,1)$ have at least one of the first $s-1$ components nonzero, the double sum at the r.h.s. of the last equality is not lesser than $\sum_{\vartheta\in P(s)} \mathcal{G}_1^{\vartheta_1} \cdots \mathcal{G}_s^{\vartheta_s} - \mathcal{G}_s$, so that

$$2\sum_{j=1}^{s-1} \mathcal{G}_j \mathcal{G}_{s-j} \geq \mathcal{F}_s + \sum_{\vartheta\in P(s)} \mathcal{G}_1^{\vartheta_1} \cdots \mathcal{G}_s^{\vartheta_s} - \mathcal{G}_s = \mathcal{G}_s$$

where the last equality follows from (22). ■

It only remains to obtain an explicite estimate for the numbers $\mathcal{G}_j$ using (21). This can be done in a very efficient way by using a procedure, rather common in numerical analysis, which consists in confrontig the $\mathcal{G}_j$ with the Taylor coefficients of a suitable analytic function.

Consider the function of the complex variable z

$$W(z) = \frac{1}{2}\left[1 - \sqrt{1-8z}\right] = 2\sum_{s=1}^{\infty} W_s z^s ,$$

which is analytic for $|z| < 1/8$ and satisfies

(24) $\quad W(z) = 2z + [W(z)]^2$

It is easy to see that this equality implies that the coefficients $2W_s$ of the power series for $W(z)$ have to be related by

$$(25)\qquad \begin{cases} W_1 = 1 \\ W_s = 2 \sum_{j=1}^{s-1} W_j \, W_{s-j} & (s \geq 2) \end{cases}$$

On the other side, using (25) one easily verifies that

$$\frac{dW}{dz}(z) = 2[1-2W(z)]^{-1}$$

$$\frac{d^s W}{dz^s}(z) = (2s-3)!!\; 2^{2s-1}\; [1-2W(z)]^{-(2s-1)} \qquad (s \geq 2)$$

where !! denotes the bifactorial, i.e. $n!!=n(n-2)(n-4)\ldots.$ Thus, being $W(0)=0$ and $(2s-3)!! \leq (2s-2)!! = (s-1)!2^{s-1}$, it follows that, for all $s \geq 1$:

$$(26)\qquad W_s = \frac{1}{s!\,2} \frac{d^s W}{dz^s}(0) \leq \frac{1}{s}\, 8^{s-1} \leq 8^{s-1}$$

By confronting (25) with (21) one then sees that, for all $s \geq 1$:

$$(27)\qquad \mathcal{G}_s \leq 8^{s-1}$$

At this point the proof of proposition A.3 is complete since (17) reduces to (7) with $A=8B$. To obtain (9) simply note that, since $\delta \xi < 1$, one has $M \|f\|_D < 8B$ so that $\|\mathcal{X}_s\|_{D'_s} \leq M \|g_s\|_{D_s} \leq (8B)^{s-1} M \|f\|_D \leq (8B)^s = A^s$.

A.5 PROOF OF PROPOSITION 2.1

It follows from proposition A.3 that, since $D'_r \supset D-(\delta,\xi)$, the generating functions χ_s are analytic on $D-(\delta,\xi)$ and satisfy $\|\chi_s\|_{D-(\delta,\xi)} \le A^s$. Recalling proposition 1.2 and the definitions (8) and (2) of A and M we get that, for all $|\varepsilon| < \varepsilon^*_P$, with

$$\varepsilon^*_P = \frac{\delta\xi}{2^5 A} = \frac{\gamma\,\delta^2\,\xi^{2(n+1)}}{2^{12+5n}\,n^n\,r^{2(n+1)}\,\|f\|_D} \equiv \varepsilon^*_r$$

$\phi_\varepsilon \equiv \phi_\varepsilon^{(\chi_1,\dots,\chi_r)}$ is an analytic canonical diffeomorphism of $D-(2\delta,2\xi)$ onto its image which is $(|\varepsilon|/4\varepsilon^*_P)(\delta,\xi)$-close to the identity. By the very definition of the χ_s in propositions A.1 and A.3, ϕ_ε brings $H_\varepsilon = h+\varepsilon f$ into its normal form up to order r. Furthermore, for the (r+1)-th remainder of $\phi_\varepsilon^*(h+\varepsilon f)$ we have, from (1.4.10), for all $|\varepsilon| < \varepsilon^*_P \equiv \varepsilon^*_r$:

$$\|R(\varepsilon,r+1)[h+\varepsilon f]\|_{D-2(\delta,\xi)} \le \|R(\varepsilon,r+1)h\|_{D-2(\delta,\xi)} + |\varepsilon|\,\|R(\varepsilon,r)f\|_{D-2(\delta,\xi)}$$

$$\le 2\,\|h\|_D \left(\frac{|\varepsilon|}{2\varepsilon^*_r}\right)^{r+1} + 2|\varepsilon|\,\|f\|_D \left(\frac{|\varepsilon|}{2\varepsilon^*_r}\right)^{r}$$

$$\le [\,\|h\|_D + \varepsilon^*_r\,\|f\|_D\,] \left(\frac{|\varepsilon|}{\varepsilon^*_r}\right)^{r}$$

Finally, note that each generating function χ_s is (by construction) real for real values of (p,q), and such is then also the flow $\phi_\varepsilon^{(\chi_1,\dots,\chi_r)}$, for real ε.

APPENDIX B

PROOF OF PROPOSITION 2.3

The present proof of proposition 2.3 is very similar to that of the previous Appendix. Although the different nonresonance condition and the presence of the cutoff require some modifications, the structure of the proof is the same.

Let us collect the notations and the hypotheses to be used. We write $D = D_{\varrho,\sigma}(B)$, with $B \subset R^n$ a bounded domain and $\varrho, \sigma > 0$; δ and ξ are positive numbers such that $\delta < \min(\varrho, 1)$, $\xi < \min(\sigma, 1)$; r is any positive integer. At difference with the Appendix A we now put

$$D_1 = D-(0, \xi/2)$$

$$D_j = D_1 - \frac{j-1}{r}(\delta, \xi/2) \qquad (1 \le s \le r+1)$$

$$D_j' = D_j - \frac{1}{2r}(\delta, \xi/2) \qquad (1 \le s \le r)$$

Note that $D_{r+1} = D_1-(\delta,\xi/2) = D-(\delta,\xi)$. The following hypothesis will be used:

(HB1) h is analytic on $D_\varrho(B)$ and there exist a lattice $\mathcal{M}$ of dimension $0 \le m < n$, a positive integer K and a positive number $\alpha_{r,K}$ such that, for all $p \in D_\varrho(B)$, it is $|k \cdot \omega(p)| \ge \alpha_{r,K}$ for all $k \notin \mathcal{M}$, $|k| < rK$, where $\omega(p) = \frac{\partial h}{\partial p}(p)$.

We first give an analogue of proposition A.1

Proposition B.1 Let h be as in (HB1). Assume that g_s is analytic on D_s, for some $1 \le s \le r$, and that $\tilde{g}_{s,k} = 0$ for all $|k| > sK$. Then the equation $L_h \chi_s = (1 - \Pi_h) g_s$ has a solution χ_s which is analytic on D_s', satisfies $\Pi_h \chi_s = 0$ and

(1) $\quad \| \chi_s \|_{D_s'} \le M \| g_s \|_{D_s}$

with

(2) $\quad M = \alpha_{r,K} \left(\frac{2^5 r}{\xi} \right)^n$

Moreover,

(3) $\quad \| L_h \chi_s \|_{D_s'} \le 2 \| g_s \|_{D_s}$

Proof Define $\chi_s(p,q) = \sum_{k \notin \Lambda} [ik \cdot \omega(p)]^{-1} \tilde{g}_{s,k}(p) \exp(ik \cdot q)$. Proceeding as in the proof of proposition A.1 and using (HB1) we get

$$\| \chi_s \|_{D_s'} \le \alpha_{r,K} \| g_s \|_{D_s} \sum_{k \notin \Lambda} e^{-|k| \frac{\xi}{8r}}$$

The same procedure used in the proof of lemma 2.1 then gives

$$\sum_{k \notin \Lambda} e^{-|k| \frac{\xi}{8r}} \le \sum_{\substack{k \in \mathbb{Z}^n \\ k \neq 0}} e^{-\frac{|k| \xi}{8r}} \le \left(\frac{2^5 r}{\xi} \right)^n$$

To prove (3) note that, since $\Pi_h g$ is the averaging of g_s over some (n−m)-dimensional torus, one has $\| \Pi_h g_s \|_{D_s} \le \| g_s \|_{D_s}$. Thus $\quad \| L_h \chi_s \|_{D_s'} \le \| g_s - \Pi_h g_s \|_{D_s} \le 2 \| g_s \|_{D_s}$ ∎

Remark It would be possible to perform the estimate of $\mathcal{X}_s$ also using explicitely the cutoff. Since $\tilde{g}_{s,k} = 0$ for $|k| > sK$, it turns out that $\mathcal{X}_s$ is analytic on D_s and an estimate of the type $\|\mathcal{X}_s\|_{D_s} \leq [2r(K+1)]^n \|g_s\|_{D_s}$ could be proven. Although this procedure avoids the introduction of the awful factor ξ^{-n} in the convergence radius $\varepsilon^*_{r,K}$ of the normalizing canonical transformation, it has the disadvantage of introducing in it a factor K^n; powers of K (and r) in $\varepsilon^*_{r,K}$ make worse the exponent in the final Nekhoroshev-like estimate of the remainder.

We can now proceed (almost) exactly as in Appendix A. Proposition A.2 obviously holds, as it does not depend on h at all, and we can use it now to recursively estimate the $(f^{\leq K})_{(s)}$ (with the caution that ξ has to be replaced with $\xi/2$). Moreover, there is a strict analogue of proposition A.3:

Proposition B.2 Assume that h is as in (HB1) and f is analytic on D. Let r and K be positive integers. Then it is possible to define the functions $\mathcal{X}_1, \ldots, \mathcal{X}_r$ through the recursive scheme

$$\mathcal{X}_s(p,q) = \sum_{k \notin \mathcal{M}} \frac{\tilde{g}_{s,k}(p)}{i\,k\cdot\omega(p)}\, e^{i k\cdot q}$$

$$g_1 = f$$

$$g_s = f^{\leq K}_{(s-1)} + [h_{(s)} - L_s h] \qquad (1<s\leq r)$$

with $L_s = L_{\mathcal{X}_s}$, $f^{\leq K}_{(s)} = (f^{\leq K})_{(s)} = \sum_{\nu\in P(s)} \frac{L_s^{\nu_s}\cdots L_1^{\nu_1}}{\nu_s!\cdots\nu_1!} f^{\leq K}$ and similarly for $h_{(s)}$. Moreover, each $\mathcal{X}_s$ is analytic on D' and satisfies

(4) $\|\mathcal{X}_s\|_{D'_s} \leq \tilde{A}^s$

where, with M defined by (2),

$$\tilde{A} = \frac{2^8 r^2}{\delta \xi} M \| f^{\leq K} \|_{D_1} \tag{5}$$

The proof of this proposition is just the same as that of proposition A.3 (provided that ξ is replaced with $\xi/2$ and f by $f^{\leq K}$, that M, D_s, D_s' are interpreted as those we are now using and it is used the fact, mentioned in section 2.4, that it is $\tilde{g}_{s,k} = 0$ for $|k| > sK$ for each $1 \leq s \leq r$). The reason is that in the proof of proposition A.3 the assumptions on h and the properties of the $\mathcal{X}_s$ do not enter directly, but only through the results of the propositions A.1 and A.2, of which we have here the counterparts.

We have now to explicitely estimate $\| f^{\leq K} \|_{D_1}$. To avoid the introduction of powers of K we use the fact that $D_1 = D-(0, \xi/2)$ and the exponential decay of the $\tilde{f}_k$:

$$\begin{aligned} \| f^{\leq K} \|_{D-(0,\xi/2)} &\leq \sum_{|k| \leq K} \| \tilde{f}_k \|_{D-(0,\xi/2)} \, e^{|k|(\sigma - \xi/2)} \leq \\ &\leq \sum_{k \in \mathbb{Z}^m} \| \tilde{f}_k \|_D \, e^{|k|(\sigma - \xi/2)} \leq \\ &\leq \| f \|_D \sum_{k \in \mathbb{Z}^m} e^{-|k| \xi/2} \leq \| f \|_D \left(\frac{8}{\xi} \right)^n \end{aligned} \tag{6}$$

It then follows from (2), (4) and (5) that, for each $s=1,\dots,r$, one has

$$\| \mathcal{X}_s \|_{D_s'} \leq A^s \tag{7}$$

with

$$(8) \qquad A = \frac{2^{8(n+1)} \, r^{n+2} \, \alpha_{r,k}}{\delta \, \xi^{2n+1}} \, \|f\|_D$$

At this point, by invoking proposition 1.2 we have that, for all $|\varepsilon| < \varepsilon^*_{r,k}$, with

$$(9) \qquad \varepsilon^*_{r,k} = \frac{\delta \xi}{2^6 A} = \frac{\delta^2 \, \xi^{2(n+1)}}{2^{8n+14} \, r^{n+2} \, \alpha_{r,k} \, \|f\|_D} ,$$

$\phi_\varepsilon^{(X_1,..,X_r)}$ is an analytic canonical diffeomorphism $\frac{\varepsilon}{4\varepsilon^*_{r,k}}(\delta,\xi)$-close to the identity of $D-2(\delta,\xi)$ onto its image. Moreover, because of the very construction of the X_s :

$$\phi_\varepsilon^{(X_1,..,X_r)*}[h+\varepsilon f^{\leq k}] = h + \sum_{s=1}^{r} \varepsilon^s Z_s + R(\varepsilon,r+1)[h+\varepsilon f^{\leq k}]$$

with $\Pi_h Z_s = 0$. Denoting $\phi_\varepsilon = \phi_\varepsilon^{(X_1,\dots,X_r)}$ we then have on $D-2(\delta,\xi)$:

$$\phi_\varepsilon^* [h+\varepsilon f] = h + \sum_{s=1}^{r} \varepsilon^s Z_s + R_{r,k}(h+\varepsilon f)$$

with

$$R_{r,k}(h+\varepsilon f) = R(\varepsilon,r+1)[h+\varepsilon f^{\leq k}] + \varepsilon \phi_\varepsilon^* f^{>k}$$

To complete the proof of proposition 2.3 it only remains to estimate this remainder. Using proposition 1.2 we have, for all $|\varepsilon| < \varepsilon^*_{r,k}$ and all $r \geq 1$ (with $H_\varepsilon = h+\varepsilon f$):

$$\| R_{r,K} (H_\varepsilon) \|_{D-2(\delta,\xi)} \leq \| h \|_D \left(\frac{|\varepsilon|}{\varepsilon^*_{r,K}} \right)^{r+1} + |\varepsilon| \, \| f^{\leq K} \|_{D-(\delta,\xi)} \left(\frac{|\varepsilon|}{\varepsilon^*_{r,K}} \right)^{r}$$

$$+ |\varepsilon| \, \| f^{>K} \|_{D-2(\delta,\xi)}$$

and using (5) (with ξ in place of $\xi/2$) and lemma 2.1:

$$\| R_{r,K} (H_\varepsilon) \|_{D-2(\delta,\xi)} \leq \| h \|_D \left(\frac{|\varepsilon|}{\varepsilon^*_{r,K}} \right)^{r+1} + |\varepsilon| \, \| f \|_D \left(\frac{4}{\xi} \right)^{n} +$$

$$+ |\varepsilon| \left(\frac{8}{\xi} \right)^{n} \| f \|_D \; e^{-K\xi/2}$$

from which (2.4.5) immediately follows.

REFERENCES

[1] R. Abraham and J.E. Marsden: Foundations of mechanics. Benjamin, Reading, Mass. (1978)

[2] V.I. Arnold: Small denominators and problems of stability of motion in classical and celestial mechanics. Usp. Mat. Nauk. 18,91 (1963) [Russ. Math. Surv. 18,85-191 (1963)]

[3] V.I. Arnold: Instability of dynamical systems with several degrees of freedom. Dokl. Akad. Nauk. SSSR 156,9 (1964) [Sov. Math. Dokl. 6,581-585 (1964)]

[4] V.I. Arnold: Mathematical methods of classical mechanics. Springer-Verlag, New York (1978)

[5] V.I.Arnold: Geometrical methods in the theory of ordinary differenyial equations. Springer-Verlag, New York (1983)

[6] V.I. Arnold and A. Avez: Problems ergodiques de la mecanique classique. Gauthier-Villars, Paris (1967)

[7] G. Benettin, L. Galgani and A. Giorgilli: A proof of Kolmogorov theorem on invariant tori using canonical transformations defined by the Lie method. Nuovo Cimento 79B,201-223 (1984)

[8] G. Benettin, L. Galgani and A. Giorgilli: Classical perturbation theory for system of weakly coupled rotators. Nuovo Cimento, 89B,89-102 (1985)

[9] G. Benettin, L. Galgani and A. Giorgilli: A proof of Nekhoroshev's theorem for the stability times in nearly integrable Hamiltonian system. Cel. Mech. 37,1-25 (1985)

[10] G. Benettin, L. Galgani and A. Giorgilli: Poincare's non-existence theorem and classical perturbation theory for nearly integrable Hamiltonian systems. In R. Livi and A. Politi (eds): Advances in nonlinear dynamics and stochastic processes, World Scientific, Singapore (1985)

[11] G. Benettin and G.Gallavotti: Stability of motions near resonances in quasi integrable Hamiltonian systems. J. Stat. Phys. 44,293-338 (1986)

[12] J.A. Campbell and W.H. Jefferys: Equivalence of the perturbation theories of Hori and Deprit. Cel. Mech.2,467-473 (1970)

[13] J.R. Cary: Lie transform perturbation theory for Hamiltonian systems, Phys. Rep. 79,129-151 (1981)

[14] J.S. Choi and B.D. Tapley: An extended canonical perturbation method. Cel. Mech.7,77-90 (1973)

[15] A. Deprit: Canonical transformations depending on a small parameter. Cel. Mech. 1,12-30 (1969)

[16] A. Deprit, J. Henrard, J.F. Price and A. Rom: Birkhoff's normalization. Cel. Mech. 1,222-251 (1969)

[17] A. Dragt and J.M. Finn: Lie series and invariant functions for analytic symplectic maps. J. Math. Phys.17,2215 (1969)

[18] J.M. Finn: Lie transforms: a perspective. In A.W. Saenz, W.W. Zachary and R. Cawley (eds): Local and global methods of nonlinear dynamics. Lect. Not. Phys. 252, Springer-Verlag, Berlin (1986), pg. 63-86

[19] G. Gallavotti: Some problems in Hamiltonian perturbation theory. Rendiconti del seminario matematico di Torino (1982)

[20] G.E.O. Giacaglia: Perturbation methods in non-linear systems. Springer-Verlag, New York (1972)

[21] A. Giorgilli and L. Galgani: Formal integrals for an autonomous Hamiltonian system near an equilibrium point. Cel. Mech.17,267-280 (1978)

[22] A. Giorgilli and L. Galgani: Rigorous estimates for the series expansion of Hamiltonian perturbation theory. Cel. Mech. 37, 95-112 (1985)

[23] E.A. Grebenikov and Yu.A. Ryabov: Constructive methods in the analysis of nonlinear systems. MIR, Moskow (1983)

[24] W. Grobner: Serie di Lie e loro applicazioni. Cremonese, Roma (1973)

[25] J. Henrard: On a perturbation theory using Lie transforms. Cel. Mech. 3, 107-120 (1970)

[26] J. Henrard: The algorithm of the inverse for Lie transform. in B.D. Tapley and V. Szebehely (eds), Recent advances in dynamical astronomy, Reidel, Dordrecht (1973), pg. 250-259

[27] J. Henrard and J. Roels: Equivalence for Lie transforms. Cel. Mech.10, 497-512 (1974)

[28] E. Hille: Ordinary differential equations in the complex domain. Wiley and Sons, New York (1973)

[29] G. Hori: Theory of general perturbations with unspecified canonical variables. Publ. Astron. Soc. Japan 18, 287-296 (1966)

[30] E.L. Ince: Ordinary differential equations. Dover, New York, N.Y. (1956)

[31] A.A. Kamel: Expansion formulae in canonical transformations depending on a small parameter. Cel. Mech.,1,190-199 (1969)

[32] A.A. Kamel: Perturbation method in the theory of nonlinear oscillations. Cel. Mech. 3,90-106 (1970)

[33] A.A. Kamel: Note on the evolution of the Lie-Deprit transform. In A.W. Saenz, W.W. Zachary and R. Cawley (eds): Local and global methods of nonlinear dynamics. Lect. Not. Phys. 252, Springer-Verlag, Berlin (1986), pg. 49-62

[34] S. Lie: Theorie der Transformationgruppen. Teubner, Leipzig (1888)

[35] F. Magri and L. Mistrangioli: Su una estensione della teoria delle serie di Lie. Preprint (1986)

[36] W.A. Mersman: Explicit recursive algorithms for the construction of equivalent canonical transformations. Cel. Mech. 3, 384-389 (1971)

[37] J. Moser: Stable and random motions in dynamical systems. Princeton Univ. Press (1973)

[38] N.N. Nekhoroshev: An exponential estimate of the time of stability of nearly-integrable Hamiltonia systems. Usp. Mat. Nauk. 32(6), 5-66 (1977) [Russ. Math. Surv. 32(6), 1-65 (1977)]

[39] N.N. Nekhoroshev: An exponential estimate of the time of stability of nearly integrable Hamiltonian systems. II. Trudy Sem. Petrows 5, 5-50 (1979), translated in O.A. Oleinik (ed), Topics in modern mathematics: Petrovkii

Seminar no.5. Consultants Bureau, New York (1985)

[40] H. Shniad: The equivalence of Von Zeipel mappings and Lie transforms. Cel. Mech. 2, 114-120 (1970)