

# IMPROVED CONVERGENCE ESTIMATES FOR THE SCHRÖDER-SIEGEL PROBLEM

*ANTONIO GIORGILLI*

*Dipartimento di Matematica, Università degli Studi di Milano,  
Via Saldini 50, 20133 — Milano, Italy.*

*UGO LOCATELLI*

*Dipartimento di Matematica, Università degli Studi di Roma “Tor Vergata”,  
Via della Ricerca Scientifica 1, 00133 — Roma, Italy.*

*MARCO SANSOTTERA*

*Università degli Studi di Milano,  
Via Saldini 50, 20133 — Milano, Italy,  
Département de Mathématique, University of Namur & NAXYS,  
Rempart de la Vierge 8, B-5000 — Namur, Belgium.*

**Abstract.** We reconsider the Schröder-Siegel problem of conjugating an analytic map in  $\mathbb{C}$  in the neighborhood of a fixed point to its linear part, extending it to the case of dimension  $n > 1$ . Assuming a condition which is equivalent to Bruno’s one on the eigenvalues  $\lambda_1, \dots, \lambda_n$  of the linear part we show that the convergence radius  $\varrho$  of the conjugating transformation satisfies  $\ln \varrho(\lambda) \geq -C\Gamma(\lambda) + C'$  with  $\Gamma(\lambda)$  characterizing the eigenvalues  $\lambda$ , a constant  $C'$  not depending on  $\lambda$  and  $C = 1$ . This improves the previous results for  $n > 1$ , where the known proofs give  $C = 2$ . We also recall that  $C = 1$  is known to be the optimal value for  $n = 1$ .

## 1. Introduction and statement of the result

We reconsider the classical problem of iteration of analytic functions, previously investigated by Schröder [14] and Siegel [15], extending it to higher dimension. Our aim is to improve the existing results on the convergence radius of an analytic near the identity transformation that conjugates the map to its linear part, thus producing general and possibly optimal estimates.

We consider an analytic map of the complex space  $\mathbb{C}^n$  into itself that leaves the origin fixed, so that it may be written as

$$(1) \quad x' = \Lambda x + v_1(x) + v_2(x) + \dots, \quad x \in \mathbb{C}^n$$

where  $\Lambda$  is a  $n \times n$  complex matrix and  $v_s(x)$  is a homogeneous polynomial of degree  $s + 1$ . The series is assumed to be convergent in a neighborhood of the origin of  $\mathbb{C}^n$ . We also assume that  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$  is a diagonal complex matrix with non resonant eigenvalues, in a sense to be clarified in a short.

The problem raised by Schröder is the following: to find an analytic near the identity coordinate transformation written as a convergent power series

$$(2) \quad y_j = x_j + \varphi_{1,j}(x) + \varphi_{2,j}(x) + \dots, \quad j = 1, \dots, n,$$

with  $\varphi_{s,j}$  of degree  $s + 1$  which conjugates the map (1) to its linear part, namely in the new coordinates  $y$  the map takes the form

$$(3) \quad y' = \Lambda y.$$

Let us write the eigenvalues of the complex matrix  $\Lambda$  in exponential form, namely as  $\lambda_j = e^{\mu_j + i\omega_j}$ , with real  $\mu_j$  and  $\omega_j$ . We define the sequence  $\{\beta_r\}_{r \geq 0}$  of positive real numbers as

$$(4) \quad \beta_0 = 1, \quad \beta_r = \min_{|k|=r+1} |e^{\langle k, \mu + i\omega \rangle - \mu_j - i\omega_j} - 1|, \quad r \geq 1$$

The eigenvalues are said to be non resonant if  $\beta_r \neq 0$  for  $r \geq 1$ . This is enough in order to positively answer the question raised by Schröder *in formal sense*, i.e., disregarding the question of convergence.

We briefly recall some known results that motivate the present work. For a detailed discussion of the problem we refer to Arnold's book [1]. The eigenvalues are said to belong to the *Poincaré domain* in case all  $\mu_j$ 's have the same sign (i.e., the eigenvalues  $\lambda$  are all inside or all outside the unit circle in the complex plane). The complement of the Poincaré domain is named *Siegel domain*. In the latter case the problem raised by Schröder represented a paradigmatic case of perturbation series involving small divisors, since the sequence  $\{\beta_r\}_{r \geq 0}$  above has zero as inferior limit.

The formal solution of the problem has been established by Schröder [14] in 1871 for the case  $n = 1$ , when the Siegel domain is the unit circle, i.e.,  $\lambda = e^{i\omega}$ . The problem of convergence in this case has been solved by Siegel [15] in 1942, assuming a strong non resonance condition of diophantine type. Siegel's proof is worked out using the classical method of majorants introduced by Cauchy, and makes use of a delicate number-theoretical lemma, often called Siegel's lemma, putting particular emphasis on the relevance of Diophantine approximations. This was actually the first proof of convergence of a meaningful problem involving small divisors.

A condition of diophantine type was also used by Kolmogorov in his proof of persistence of invariant tori under small perturbation of integrable Hamiltonian systems, and has become a standard one in KAM theory. It should be noted that Kolmogorov introduced also an iteration scheme exhibiting a fast convergence, that he described as "similar to Newton's method" and is often referred to as "quadratic method". The fast convergence scheme applies also to the case of maps discussed here: see, e.g., [1], § 28.

In the last decades two problems have been raised (among many others). The first one is concerned with determining the optimal non resonance condition; the second one with the size of the convergence radius of the transformation to normal form. In the case  $n = 1$  both questions have been answered exploiting the geometric renormalization approach introduced by Yoccoz (see [16], [17] [2]). The optimal set of rotation numbers for which an analytic transformation to linear normal form exists has been identified

with the set of Bruno numbers. These numbers obey a non resonance condition weaker than the diophantine one introduced by Siegel.

Bruno's condition in case  $n > 1$  may be stated as follows. Let us introduce the sequence  $\{\alpha_r\}_{r \geq 0}$  defined as

$$(5) \quad \alpha_r = \min_{0 \leq s \leq r} \beta_s, \quad r \geq 0.$$

The latter sequence is monotonically non increasing, and if the eigenvalues belong to Siegel's domain has zero as lower limit. The non resonance condition is

$$-\sum_{k>0} \frac{\ln \alpha_{2^k-1}}{2^k} = B < \infty.$$

The series expansion of the transformation that solves Schröder's problem is proved to be convergent in a disk  $\Delta_\varrho$  centered at the origin with radius  $\varrho(\lambda)$  satisfying

$$(6) \quad \ln \varrho \geq -CB + C',$$

where  $C'$  is a constant independent of  $\lambda$ . For  $n = 1$  the optimal value has been proved to be  $C = 1$ . However, the geometric renormalization methods can not be extended to the case  $n > 1$ , for which only the value  $C = 2$  has been found. A proof that gives  $C = 1$  in the case  $n = 1$  using the majorant method has been given in [3], but has not been extended to the case  $n > 1$ .

In this paper we prove that the same bound with  $C = 1$  holds true in any finite dimension. We obtain this result by implementing a representation of the map with the method of Lie series and Lie transforms, and producing convergence estimates in the spirit of Cauchy's majorants method. The formal implementation of the representation of maps may be found in [11].

This paper extends to the case of maps a previous similar result for the case of the Poincaré-Siegel theoretical center problem, where the problem of linearization of an analytic system of differential equations in the neighborhood of a fixed point is considered. We stress that the interest of the method is not limited to the cases mentioned here. E.g.: for applications to KAM theory see [4], [5] and [12]; a proof of Lyapounov's theorem on the existence of periodic orbits in the neighborhood of the equilibrium is given in [10]; extensions of Lyapounov's theorem may be found in [6] and [8].

We come now to a formal statement of our result. We assume the following

**Condition  $\tau$ :** The sequence  $\alpha_r$  above satisfies

$$(7) \quad -\sum_{r \geq 1} \frac{\ln \alpha_r}{r(r+1)} = \Gamma < \infty.$$

**Theorem 1:** Consider the map (1) and assume that the eigenvalues of  $\Lambda$  are non-resonant and satisfy condition  $\tau$ . Then there exists a near to identity coordinate transformation  $y = x + \psi(x)$ , with  $\psi$  analytic at least in the polydisk of radius  $B^{-1}e^{-\Gamma}$ , where  $B > 0$  is a universal constant, which transforms the map into the normal form  $y' = \Lambda y$ .

Our condition  $\tau$  is equivalent to Bruno's one. We have indeed

$$\Gamma \leq B \leq 2\Gamma .$$

However, our formulation of condition  $\tau$  comes out naturally from our analysis of accumulation of small divisors and turns out to be the key that allows us to find the estimate with  $C = 1$ . For a brief discussion of the relations between the two conditions see [9].

The paper is organized as follows. In section 2 we include the essential information on the formal algorithm, referring to [11] for details. In section 3 we give the technical estimates that lead to the proof of convergence of the formal algorithm. In particular we include a complete discussion of the mechanism of accumulation of small divisors which allows us to have an accurate control. We remark in particular that we do not need to use the Siegel lemma, because the latter controls combinations of small divisors that do not occur in our scheme. The estimates of section 3 are used in section 4 in order to complete the proof of the main theorem. A technical appendix follows.

## 2. Formal algorithm

We represent a map of type (1) using Lie transforms. A detailed exposition of the representation algorithm and of the method of reduction to normal form is given in [11], where the problem is treated at a formal level. We refer to that paper for details concerning the formal setting, while in this paper we pay particular attention to the problem of convergence. Here we just include the definitions that we shall need later, so that also the notations will be fixed. Some relevant lemmas are reported in appendix A.

We start with the definition of Lie series and Lie transform. Let  $X_s(x)$  be a vector field on  $\mathbb{C}^n$  whose components are homogeneous polynomials of degree  $s + 1$ . We shall say that  $X_s(x)$  is of *order*  $s$ , as indicated by the label. Moreover, in the following we shall denote by  $X_{s,j}$  the  $j$ -th component of the vector field  $X_s$ . The *Lie series* operator is defined as

$$(8) \quad \exp(L_{X_s}) = \sum_{j \geq 0} \frac{1}{j!} L_{X_s}^j$$

where  $L_{X_s}$  is the Lie derivative with respect to the vector field  $X_s$ .

Let now  $X = \{X_j\}_{j \geq 1}$  be a sequence of polynomial vector fields of degree  $j + 1$ . The *Lie transform* operator  $T_X$  is defined as

$$(9) \quad T_X = \sum_{s \geq 0} E_s^X ,$$

where the sequence  $E_s^X$  of linear operators is recursively defined as

$$(10) \quad E_0^X = 1 , \quad E_s^X = \sum_{j=1}^s \frac{j}{s} L_{X_j} E_{s-j}^X .$$

The superscript in  $E^X$  is introduced in order to specify which sequence of vector fields is intended. By letting the sequence to have only one vector field different from zero, e.g.,  $X = \{0, \dots, 0, X_k, 0, \dots\}$  it is easily seen that one gets  $T_X = \exp(L_{X_k})$ .

## 2.1 Representation and conjugation of maps

We recall the representation of maps introduced in [11] together with some formal results that we are going to use here. Let  $\Lambda = e^A$ , i.e., in our case,  $A = \text{diag}(\mu_1 + i\omega_1, \dots, \mu_n + i\omega_n)$  with  $\lambda_j = e^{\mu_j + i\omega_j}$ . Remark that we may express the linear part of the map as a Lie series by introducing the exponential operator  $R = \exp(L_{Ax})$ . The action of the operator  $R$  on a function  $f$  or on a vector field  $V$  is easily calculated as

$$(11) \quad (\mathbf{R}f)(x) = f(\Lambda x) , \quad (\mathbf{R}V)(x) = \Lambda^{-1}V(\Lambda x) .$$

The first result is concerned with the representation of the map (1) using a Lie transform.

**Lemma 1:** *There exist generating sequences of vector fields  $V = \{V_s(x)\}_{s \geq 1}$  and  $W = \{W_s(x)\}_{s \geq 1}$  with  $W_s = \mathbf{R}V_s$  such that the map (1) is represented as*

$$(12) \quad x' = \mathbf{R} \circ T_V x \quad \text{and} \quad x' = T_W \circ \mathbf{R} x$$

The second result is concerned with the composition of Lie transforms.

**Lemma 2:** *Let  $X, Y$  be generating sequences. Then one has  $T_X \circ T_Y = T_Z$  where  $Z$  is the generating sequence recursively defined as*

$$(13) \quad Z_1 = X_1 + Y_1 , \quad Z_s = X_s + Y_s + \sum_{j=1}^{s-1} \frac{j}{s} E_{s-j}^X Y_j .$$

The latter formula reminds the well known Baker-Campbell-Hausdorff composition of exponentials. The difference is that the result is expressed as a Lie transform instead of an exponential, which makes the formula more effective for our purposes.

The third result gives the algorithm that we shall use in order to conjugate the map (1) to its linear part. We formulate it in a more general form, looking for conjugation between maps. Let two maps

$$(14) \quad x' = T_W \circ \mathbf{R} x , \quad y' = T_Z \circ \mathbf{R} y$$

be given, where  $W = \{W_s\}_{s \geq 1}$ ,  $Z = \{Z_s\}_{s \geq 1}$  are generating sequences. We say that the maps are conjugated up to order  $r$  in case there exists a finite generating sequence  $X = \{X_1, \dots, X_r\}$  such that the transformation  $y = S_X^{(r)} x$  makes the difference between the maps to be of order higher than  $r$ , i.e.,

$$S_X^{(r)} x' \Big|_{x' = T_W \circ \mathbf{R} x} - T_Z \circ \mathbf{R} y \Big|_{y = S_X^{(r)} x} = \mathcal{O}(r+1) ,$$

where  $S_X^{(r)} = \exp(L_{X_r}) \circ \dots \circ \exp(L_{X_1})$ . Suppose that we have  $W_1 = Z_1, \dots, W_r = Z_r$ . Then the maps are conjugated up to order  $r$ , since one has  $T_W x - T_Z x = \mathcal{O}(r+1)$ .

**Lemma 3:** *Let the generating sequences of the maps (14) coincide up to order  $r - 1$  and let  $X_r$  be a vector field of order  $r$  generating the near the identity transformation  $y = \exp(L_{X_r}x)$ . Then the maps are conjugated up to order  $r$  if*

$$(15) \quad T_Z = \exp(L_{X_r}) \circ T_W \circ \exp(L_{-RX_r}) .$$

The vector field  $X_r$  must satisfy the equation

$$(16) \quad DX_r = W_r - Z_r , \quad D = R - 1 .$$

## 2.2 Construction of the normal form

Following Schröder and Siegel, we want to conjugate the map (1) to its linear part. That is: writing the map as

$$(17) \quad x' = T_{W^{(0)}} \circ Rx$$

with a known sequence  $W^{(0)}$  of vector fields, we want to reduce it to the linear normal form

$$(18) \quad x' = Rx .$$

To this end we look for a generating sequence  $\{X_r\}_{r \geq 1}$  of vector fields and a corresponding sequence  $\{W^{(r)}\}_{r \geq 1}$  satisfying  $W_1^{(r)} = \dots = W_r^{(r)} = 0$ . We emphasize that the map  $x' = T_{W^{(r)}} \circ Rx$  is conjugated to (18) up to order  $r$ . We say that  $W^{(r)}$  is in normal form up to order  $r$ .

According to (16) we should solve for  $X_r$  the equation

$$(19) \quad DX_r = W_r^{(r-1)} , \quad D = R - 1 .$$

The operator  $D$  is diagonal on the basis of monomials  $x^k \mathbf{e}_j = x_1^{k_1} \dots x_n^{k_n} \mathbf{e}_j$ , where  $(\mathbf{e}_1, \dots, \mathbf{e}_n)$  is the canonical basis of  $\mathbb{C}^n$ . For we have

$$(20) \quad D x^k \mathbf{e}_j = (e^{\langle k, \mu + i\omega \rangle - \mu_j - i\omega_j} - 1) x^k \mathbf{e}_j .$$

Thus, provided the eigenvalues of  $D$  are not zero, the vector field  $X_r$  is determined as

$$X_r = \sum_{j=1}^n \mathbf{e}_j \sum_k \frac{w_{j,k}}{e^{\langle k, \mu + i\omega \rangle - \mu_j - i\omega_j} - 1} x^k ,$$

where  $w_{j,k}$  are the coefficients of  $W_r^{(r-1)}$ .

Next, we need an explicit form for the transformed sequence  $W^{(r)}$ . We use the conjugation formula (15) replacing  $W$  and  $Z$  with  $W^{(r-1)}$  and  $W^{(r)}$ , respectively. It is convenient to introduce an auxiliary vector field  $V^{(r)}$  and to split the formula as

$$(21) \quad \begin{aligned} T_{V^{(r)}} &= T_{W^{(r-1)}} \circ \exp(L_{-RX_r}) . \\ T_{W^{(r)}} &= \exp(L_{X_r}) \circ T_{V^{(r)}} , \end{aligned}$$

A more explicit form comes from lemma 2, recalling that  $T_X = \exp(L_{X_r})$  if one considers the generating sequence  $X = \{0, \dots, 0, X_r, 0, \dots\}$ . The auxiliary vector field  $V^{(r)}$  is

determined as

$$(22) \quad \begin{aligned} V_r^{(r)} &= W_r^{(r-1)} - \mathbf{R}X_r, \\ V_s^{(r)} &= W_s^{(r-1)} - \frac{r}{s} E_{s-r}^{(r-1)} \mathbf{R}X_r \quad \text{for } s > r, \end{aligned}$$

where we use the simplified notation  $E^{(r-1)}$  in place of  $E^{(W^{(r-1)})}$ . Having so determined the sequence  $V^{(r)}$  we calculate the transformed sequence  $W^{(r)}$  as

$$(23) \quad \begin{aligned} W_r^{(r)} &= V_r^{(r)} + X_r; \\ W_s^{(r)} &= V_s^{(r)} + \frac{1}{s} \sum_{k=1}^{\lfloor s/r \rfloor - 1} \frac{s - kr}{k!} L_{X_r}^k V_{s-kr}^{(r)}, \quad s > r. \end{aligned}$$

Here a remark is in order. The formulæ above define the sequences  $V_r^{(r)}, V_{r+1}^{(r)}, \dots$  and  $W_{r+1}^{(r)}, W_{r+2}^{(r)}, \dots$  starting with terms of order  $r$  and  $r + 1$ , respectively. This is fully natural, because all terms of lower order vanish due to  $W^{(r)}$  being in normal form up to order  $r$ . Moreover we emphasize that in view of (16) we should determine  $X_r$  by solving the equation  $\mathbf{D}X_r = W_r^{(r-1)}$ , since we want  $W_r^{(r)} = 0$ .

### 3. Quantitative estimates

Our aim now is to complete the formal algorithm of the previous section with quantitative estimates that will lead to the proof of convergence of the transformation to normal form. The main result of this section is the iteration lemma of section 3.3 below. However, we must anticipate a few technical tools.

#### 3.1 Norms on vector fields and generalized Cauchy estimates

For a homogeneous polynomial  $f(x) = \sum_{|k|=s} f_k x^k$  (using multiindex notation and with  $|k| = |k_1| + \dots + |k_n|$ ) with complex coefficients  $f_k$  and for a homogeneous polynomial vector field  $X_s = (X_{s,1}, \dots, X_{s,n})$  we use the *polynomial norm*

$$(24) \quad \|f\| = \sum_k |f_k|, \quad \|X_s\| = \sum_{j=1}^n \|X_{s,j}\|.$$

The following lemma allows us to control the norms of Lie derivatives of functions and vector fields.

**Lemma 4:** *Let  $X_r$  be a homogeneous polynomial vector field of degree  $r + 1$ . Let  $f_s$  and  $v_s$  be a homogeneous polynomial and vector field, respectively, of degree  $s + 1$ . Then we have*

$$(25) \quad \|L_{X_r} f_s\| \leq (s + 1) \|X_r\| \|f_s\| \quad \text{and} \quad \|L_{X_r} v_s\| \leq (r + s + 2) \|X_r\| \|v_s\|.$$

**Proof.** Write  $f_s = \sum_{|k|=s+1} b_k x^k$  with complex coefficients  $b_k$ . Similarly, write the  $j$ -th component of the vector field  $X_r$  as  $X_{r,j} = \sum_{|k'|=r+1} c_{j,k'} x^{k'}$ . Recalling that  $L_{X_r} f_s =$

$\sum_{j=1}^n X_{r,j} \frac{\partial f_s}{\partial x_j}$  we have

$$L_{X_r} f_s = \sum_{j=1}^n \sum_{k,k'} \frac{c_{j,k'} k_j b_k}{x_j} x^{k+k'} .$$

Thus in view of  $|k_j| \leq s+1$  we have

$$\|L_{X_r} f_s\| \leq (s+1) \sum_{j=1}^n \sum_{k'} |c_{j,k'}| \sum_k |b_k| = (s+1) \|X_r\| \|f_s\| ,$$

namely the first of (25). In order to prove the second inequality recall that the  $j$ -th component of the Lie derivative of the vector field  $v_s$  is

$$(L_{X_r} v_s)_j = \sum_{l=1}^n \left( X_{r,l} \frac{\partial v_{s,j}}{\partial x_l} - v_{s,l} \frac{\partial X_{r,j}}{\partial x_l} \right) .$$

Then using the first of (25) we have

$$\left\| \sum_{l=1}^n \left( X_{r,l} \frac{\partial v_{s,j}}{\partial x_l} - v_{s,l} \frac{\partial X_{r,j}}{\partial x_l} \right) \right\| \leq (s+1) \|X_r\| \|v_{s,j}\| + (r+1) \|v_s\| \|X_{r,j}\| ,$$

which readily gives the wanted inequality in view of the definition (24) of the polynomial norm of a vector field. Q.E.D.

**Lemma 5:** *Let  $V_r$  be a homogeneous polynomial vector field of degree  $r+1$ . Then the solution  $X_r$  of the equation  $DX_r = V_r$  satisfies*

$$(26) \quad \|X_r\| \leq \frac{1}{\alpha_r} \|V_r\| , \quad \|RX_r\| \leq \frac{1+\alpha_r}{\alpha_r} \|V_r\| ,$$

with the sequence  $\alpha_r$  defined by (5)

**Proof.** The first inequality is a straightforward consequence of the definition (24) of the norm and of the sequence  $\alpha_r$  in terms of  $\beta_r$  defined by (4). For if  $v_{j,k}$  are the coefficients of  $V_r$  then the coefficients of  $X_r$  are bounded by  $|v_{j,k}|/\beta_r \leq |v_{j,k}|/\alpha_r$ . The second inequality follows from  $RX_r = X_r + V_r$ , which gives the stated inequality. Q.E.D.

### 3.2 Accumulation of small divisors

Lemma 5 shows that solving the equation for every vector field of the generating sequence introduces a divisor  $\alpha_r$ . Such divisors do accumulate, and our aim now is to analyze in detail the process of accumulation. It will be convenient to introduce a further sequence  $\{\sigma_r\}_{r \geq 0}$  defined as

$$(27) \quad \sigma_0 = 1 , \quad \sigma_r = \frac{\alpha_r}{r^2} , \quad r \geq 1 ,$$

which will play a major role in the rest of the proof. The extra factor  $1/r^2$  can be interpreted as due to the generalized Cauchy estimates for derivatives, that are also a source of divergence in perturbation processes. The quantities  $\sigma_r$  are the actual small divisors that we must deal with. Here we follow the scheme presented in [9], with a few



*variazioni*. However, since this is a crucial part of the proof we include it in a detailed and self contained form.

The guiding remark is that the small divisors  $\sigma_r$  propagate and accumulate through the formal construction due to the use of the recursive formulæ (22) and (23). E.g., the expression  $L_{X_r} V_s^{(r)}$  will contain the product of the divisors contained in both  $X_r$  and  $V_s^{(r)}$ , with no extra divisors generated by the Lie derivative. The explicit constructive form of the algorithm allows us to have a quite precise control on the accumulation process. It is an easy remark that unfolding the recursive formulæ (22) and (23) will produce an estimate of  $\|W_s^{(r)}\|$  as a sum of many terms every one of which contains as denominator a product of  $q$  divisors of the form  $\sigma_{j_1} \cdots \sigma_{j_q}$ , with some indexes  $j_1, \dots, j_q$  and some  $q$  to be found. This is what we call the *accumulation of small divisors*, and the problem is to identify the worst product among them. The key of our argument is to focus our attention on the indexes rather than on the actual values of the divisors.

We call  $I = \{j_1, \dots, j_s\}$  with non negative integers  $j_1, \dots, j_s$  a *set of indexes*. We introduce a partial ordering as follows. Let  $I = \{j_1, \dots, j_s\}$  and  $I' = \{j'_1, \dots, j'_s\}$  be two sets of indexes with the same number  $s$  of elements. We say that  $I \triangleleft I'$  in case there is a permutation of the indexes such that the relation  $j_m \leq j'_m$  holds true for  $m = 1, \dots, s$ . If two sets of indexes contain a different number of elements we pad the shorter one with zeros and use the same definition. We also define the special sets of indexes

$$(28) \quad I_s^* = \left( \left\lfloor \frac{s}{s} \right\rfloor, \left\lfloor \frac{s}{s-1} \right\rfloor, \dots, \left\lfloor \frac{s}{2} \right\rfloor \right) .$$

**Lemma 6:** For the sets of indexes  $I_s^* = \{j_1, \dots, j_s\}$  the following statements hold true:

- (i) the maximal index is  $j_{\max} = \lfloor \frac{s}{2} \rfloor$ ;
- (ii) for every  $k \in \{1, \dots, j_{\max}\}$  the index  $k$  appears exactly  $\lfloor \frac{s}{k} \rfloor - \lfloor \frac{s}{k+1} \rfloor$  times;
- (iii) for  $0 < r \leq s$  one has

$$(\{r\} \cup I_r^* \cup I_s^*) \triangleleft I_{r+s}^* .$$

**Proof.** The claim (i) is a trivial consequence of the definition.

(ii) For each fixed value of  $s > 0$  and  $1 \leq k \leq \lfloor s/2 \rfloor$ , we have to determine the cardinality of the set  $M_{k,s} = \{m \in \mathbb{N} : 2 \leq m \leq s, \lfloor s/m \rfloor = k\}$ . For this purpose, we use the trivial inequalities

$$\left\lfloor \frac{s}{\lfloor s/k \rfloor} \right\rfloor \geq k \quad \text{and} \quad \left\lfloor \frac{s}{\lfloor s/k \rfloor + 1} \right\rfloor < k .$$

After having rewritten the same relations with  $k+1$  in place of  $k$ , one immediately realizes that a index  $m \in M_{k,s}$  if and only if  $m \leq \lfloor s/k \rfloor$  and  $m \geq \lfloor s/(k+1) \rfloor + 1$ , therefore  $\#M_{k,s} = \lfloor \frac{s}{k} \rfloor - \lfloor \frac{s}{k+1} \rfloor$ .

(iii) Since  $r \leq s$ , the definition in (28) implies that neither  $\{r\} \cup I_r^* \cup I_s^*$  nor  $I_{r+s}^*$  can include any index exceeding  $\lfloor (r+s)/2 \rfloor$ . Thus, let us define some finite sequences of non-negative integers as follows:

$$\begin{aligned} R_k &= \#\{j \in I_r^* : j \leq k\} , & S_k &= \#\{j \in I_s^* : j \leq k\} , \\ M_k &= \#\{j \in \{r\} \cup I_r^* \cup I_s^* : j \leq k\} , & N_k &= \#\{j \in I_{r+s}^* : j \leq k\} , \end{aligned}$$

where  $1 \leq k \leq \lfloor (r+s)/2 \rfloor$ . When  $k < r$ , the property (ii) of the present lemma allows us to write

$$R_k = r - \left\lfloor \frac{r}{k+1} \right\rfloor, \quad S_k = s - \left\lfloor \frac{s}{k+1} \right\rfloor, \quad N_k = r + s - \left\lfloor \frac{r+s}{k+1} \right\rfloor;$$

using the elementary estimate  $\lfloor x \rfloor + \lfloor y \rfloor \leq \lfloor x+y \rfloor$ , from the equations above it follows that  $M_k \geq N_k$  for  $1 \leq k < r$ . In the remaining cases, i.e., when  $r \leq k \leq \lfloor (r+s)/2 \rfloor$ , we have that

$$R_k = r - 1, \quad S_k = s - \left\lfloor \frac{s}{k+1} \right\rfloor, \quad N_k = r + s - \left\lfloor \frac{r+s}{k+1} \right\rfloor;$$

therefore,  $M_k = 1 + R_k + S_k \geq N_k$ . Since we have just shown that  $M_k \geq N_k \forall 1 \leq k \leq \lfloor (r+s)/2 \rfloor$ , it is now an easy matter to complete the proof. Let us first imagine to have reordered both the set of indexes  $\{r\} \cup I_r^* \cup I_s^*$  and  $I_{r+s}^*$  in increasing order; moreover, let us recall that  $\#(\{r\} \cup I_r^* \cup I_s^*) = \#I_{r+s}^* = r+s-1$ , because of the definition in (28). Thus, since  $M_1 \geq N_1$ , every element equal to 1 in  $\{r\} \cup I_r^* \cup I_s^*$  has a corresponding index in  $I_{r+s}^*$  the value of which is at least 1. Analogously, since  $M_2 \geq N_2$ , every index 2 in  $\{r\} \cup I_r^* \cup I_s^*$  has a corresponding index in  $I_{r+s}^*$  which is at least 2, and so on up to  $k = \lfloor (r+s)/2 \rfloor$ . We conclude that  $\{r\} \cup I_r^* \cup I_s^* \triangleleft I_{r+s}^*$ . *Q.E.D.*

We come now to identify the sets of indexes that describe the allowed combinations of small divisors. These are the sets

$$(29) \quad \mathcal{J}_{r,s} = \{I = \{j_1, \dots, j_{s-1}\} : j_m \in \{0, \dots, \min(r, s/2)\}, I \triangleleft I_s^*\}.$$

We shall refer to condition  $I \triangleleft I_s^*$  as the *selection rule S*. The relevant properties are stated by the following

**Lemma 7:** *For the sets of indexes  $\mathcal{J}_{r,s}$  the following statements hold true:*

- (i)  $\mathcal{J}_{r-1,s} \subset \mathcal{J}_{r,s}$ ;
- (ii) if  $I \in \mathcal{J}_{r-1,r}$  and  $I' \in \mathcal{J}_{r,s}$  then we have  $(\{r\} \cup I \cup I') \in \mathcal{J}_{r,r+s}$ .

**Remark.** Property (ii) plays a major role in controlling the accumulation of small divisors, since it gives us a control of how the indexes are accumulated. For it reflects the fact that a Lie derivative, e.g.,  $L_{X_r} V_s^{(r)}$  contains the union of the divisors in  $V_s^{(r)}$  and in  $X_r$ , taking also into account that  $X_r$  contains an extra divisor in view of lemma 5.

**Proof of lemma 7.** (i) is immediately checked in view of the definition of  $\mathcal{J}_{r,s}$ .

(ii) We have  $\#(\{r\} \cup I \cup I') = 1 + \#(I) + \#(I') = 1 + r - 1 + s - 1 = r + s - 1$ . For  $j \in \{r\} \cup I \cup I'$  we also have  $1 \leq j \leq r$  because this is true for all  $j \in I$  and for all  $j \in I'$ , and we just add an extra index  $r$ . Since  $r < s$ , we also have  $r = \min(r, (r+s)/2)$ , as required. Coming to the selection rule **S** we remark that  $\{r\} \cup I \cup I' \triangleleft \{r\} \cup I_r^* \cup I_s^*$  readily follows from  $I \triangleleft I_r^*$  and  $I' \triangleleft I_s^*$ , which are true in view of  $I \in \mathcal{J}_{r-1,r}$  and  $I' \in \mathcal{J}_{r,s}$ , so that the claim follows from property (iii) of lemma 6. *Q.E.D.*

We come now to consider the accumulation of small divisors. Recall the definition (27) of the sequence  $\sigma_r$ . We associate to the sets of indexes  $\mathcal{J}_{r,s}$  the sequence of positive real numbers  $T_{r,s}$  defined as

$$(30) \quad T_{0,s} = 1, \quad T_{r,s} = \max_{I \in \mathcal{J}_{r,s}} \prod_{j \in I} \frac{1}{\sigma_j}, \quad 0 < r \leq s.$$

**Lemma 8:** *The sequence  $T_{r,s}$  satisfies the following properties for  $1 \leq r \leq s$ :*

- (i)  $T_{r-1,s} \leq T_{r,s}$ ;
- (ii)  $\frac{1}{\sigma_r} T_{r-1,r} T_{r,s} \leq T_{r,r+s}$ .

**Proof.** (i) From property (i) of lemma 7 we readily get

$$T_{r-1,s} = \max_{I \in \mathcal{J}_{r-1,s}} \prod_{j \in I} \frac{1}{\sigma_j} \leq \max_{I \in \mathcal{J}_{r,s}} \prod_{j \in I} \frac{1}{\sigma_j},$$

since the maximum is evaluated over a larger set of indexes.

(ii) Compute

$$\begin{aligned} \frac{1}{\sigma_r} T_{r-1,r} T_{r,s} &= \frac{1}{\sigma_r} \max_{I \in \mathcal{J}_{r-1,r}} \prod_{j \in I} \frac{1}{\sigma_j} \max_{I' \in \mathcal{J}_{r,s}} \prod_{j' \in I'} \frac{1}{\sigma_{j'}} \\ &= \max_{I \in \mathcal{J}_{r-1,r}} \max_{I' \in \mathcal{J}_{r,s}} \prod_{j \in \{r\} \cup I \cup I'} \frac{1}{\sigma_j} \\ &\leq \max_{I \in \mathcal{J}_{r,r+s}} \prod_{j \in I} \frac{1}{\sigma_j} = T_{r,r+s}, \end{aligned}$$

where in the inequality of the last line property (ii) of lemma 7 has been used. *Q.E.D.*

The final estimate uses condition  $\tau$  and the definition (27) of the sequence  $\sigma_r$ .

**Lemma 9:** *Let  $\lambda$  satisfy condition  $\tau$ . Then the sequence  $T_{r,s}$  is bounded by*

$$T_{r,s} \leq \gamma^s e^{s\Gamma}, \quad \frac{1}{\sigma_s} T_{r,s} \leq \gamma^s e^{s\Gamma}$$

with some positive constant  $\gamma$  not depending on  $\lambda$ .

**Proof.** In view of  $\sigma_s \leq 1$  it is enough to prove the second inequality. We use the property (ii) of lemma 7 and the selection rule S. We readily get

$$\frac{1}{\sigma_s} T_{r,s} = \frac{1}{\sigma_s} \max_{I \in \mathcal{J}_{r,s}} \prod_{j \in I} \frac{1}{\sigma_j} \leq \max_{I \in \mathcal{J}_{r,s}} \prod_{j \in \{s\} \cup I} \frac{1}{\sigma_j} \leq \prod_{j \in \{s\} \cup I_s^*} \frac{1}{\sigma_j}.$$

By property (ii) of lemma 6 the latter product is evaluated as

$$\prod_{j \in \{s\} \cup I_s^*} \frac{1}{\sigma_j} = [\sigma_1^{q_1} \cdots \sigma_{\lfloor s/2 \rfloor}^{q_{\lfloor s/2 \rfloor}} \sigma_s^{q_s}]^{-1},$$

where  $q_k = \lfloor \frac{s}{k} \rfloor - \lfloor \frac{s}{k+1} \rfloor$  is the number of indexes in  $I_s^*$  which are equal to  $k$ . In view of the definition (27) of the sequence  $\sigma_s$  we have

$$\ln \frac{1}{\sigma_r} T_{r,s} \leq - \sum_{k=1}^s \left( \left\lfloor \frac{s}{k} \right\rfloor - \left\lfloor \frac{s}{k+1} \right\rfloor \right) (\ln \alpha_k - 2 \ln k) \leq -s \sum_{k \geq 1} \frac{\ln \alpha_k - 2 \ln k}{k(k+1)} = s(\Gamma + a),$$

where  $a = \sum_{k \geq 1} \frac{2 \ln k}{k(k+1)} < \infty$  is clearly independent of  $\lambda$ . The claim follows by just setting  $\gamma = e^a$ . *Q.E.D.*

### 3.3 Iteration lemma

This is the main lemma which allows us to control the norms of the sequence of vector fields  $\{X_r\}_{r \geq 1}$ .

**Lemma 10:** *Assume that the sequence  $W^{(0)}$  of vector fields satisfies  $\|W_s^{(0)}\| \leq \frac{C_0^{s-1}A}{s}$  with some constants  $A > 0$  and  $C_0 \geq 0$ . Then the sequence of vector fields  $\{X_r\}_{r \geq 1}$  that for every  $r$  give the normal form  $W^{(r)}$  satisfies the following estimates: there exists a bounded monotonically non decreasing sequence  $\{C_r\}_{r \geq 1}$  of positive constants, with  $C_r \rightarrow C_\infty < \infty$  for  $r \rightarrow \infty$ , such that we have*

$$(31) \quad \|X_r\| \leq T_{r-1,r} \frac{C_{r-1}^{r-1}A}{r\alpha_r}, \quad \|W_s^{(r)}\| \leq T_{r,s} \frac{C_r^{s-1}A}{s}.$$

The sequence may be recursively defined as

$$(32) \quad C_1 = 2C_0 + 16A, \quad C_r = \left(1 + \frac{1}{r^2}\right)^{1/r} \left(1 + \frac{1}{r}\right)^{1/r} C_{r-1},$$

so that one has  $C_r > 16A$ .

**Proof.** The proof proceeds by induction. For  $r = 0$  the inequality for  $W_s^{(0)}$  is nothing but the initial hypothesis, recalling that by definition we have  $T_{0,s} = 1$ . From this, the corresponding inequality for  $X_1$  immediately follows by lemma 5. The induction requires two main steps, namely: (a) estimating  $\|V_s^{(r)}\|$  as defined by (22), and (b) estimating  $\|W_s^{(r)}\|$  as defined by (23).

The estimate for  $X_r$  is determined by solving (19). In view of the induction hypothesis (31) we rewrite the claim of lemma 5 as

$$(33) \quad \|X_r\| \leq \frac{r^2 T_{r-1,r}}{\alpha_r} \cdot \frac{C_{r-1}^{r-1}A}{r^3}, \quad \|RX_r\| \leq \frac{T_{r-1,r}}{\alpha_r} \cdot \frac{(1 + \alpha_r)C_{r-1}^{r-1}A}{r}.$$

In order to estimate  $\|V_s^{(r)}\|$  we first prove that for  $s > r$  we have

$$(34) \quad \|E_{s-1}^{(1)}RX_1\| \leq T_{1,s}C_0^{s-1}A \frac{8A}{C_0} \left(1 + \frac{4A}{C_0}\right)^{s-2},$$

$$(35) \quad \|E_{s-r}^{(r-1)}RX_r\| < T_{r,s} \frac{C_{r-1}^{s-1}A}{r^2} \left(1 + \frac{1}{r}\right)^{\frac{s}{r}-2} \quad \text{for } r > 1.$$

The proof of the latter estimates is based on the general inequality

$$(36) \quad \|E_{s-r}^{(r-1)}RX_r\| \leq \sum_{k=1}^{\lfloor s/r \rfloor - 1} \Theta(r, s-r, k) T_{r-1, j_1} \cdots T_{r-1, j_k} \frac{T_{r-1,r}}{\alpha_r} \cdot \frac{(1 + \alpha_r)C_{r-1}^{s-k-1}A^{k+1}}{r}$$

where we have introduced the quantities

$$(37) \quad \Theta(r, s, k) = \sum_{\substack{j_1 + \dots + j_k = s \\ j_1, \dots, j_k \geq r}} \frac{(j_1 + \dots + j_k + r + 2)(j_1 + \dots + j_{k-1} + r + 2) \cdots (j_1 + r + 2)}{(j_1 + \dots + j_k)(j_1 + \dots + j_{k-1}) \cdots j_1}$$

The inequality (36) follows by repeatedly applying lemma 4 to the non recursive expression for  $E_{s-r}^{(r-1)}$  given in lemma 11. We describe this process in detail. First estimate

$$\|L_{W_{j_1}^{(r-1)}} \mathbf{R}X_r\| \leq (j_1 + r + 2) T_{r-1, j_1} \frac{C_{r-1}^{j_1-1} A}{j_1} \cdot \frac{T_{r-1, r}}{\alpha_r} \cdot \frac{(1 + \alpha_r) C_{r-1}^{r-1} A}{r}$$

which follows by the induction hypothesis (31), from (33) and from lemma 4. Remark that the factor  $j_1 + r + 2$  is the degree of the vector field  $L_{W_{j_1}^{(r-1)}} \mathbf{R}X_r$ . The same estimate is repeated for  $L_{W_{j_2}^{(r-1)}} L_{W_{j_1}^{(r-1)}} \mathbf{R}X_r, \dots$  thus yielding

$$\|L_{W_{j_k}^{(r-1)}} \cdots L_{W_{j_1}^{(r-1)}} \mathbf{R}X_r\| \leq \frac{(j_1 + \dots + j_k + r + 2) \cdots (j_1 + r + 2)}{j_k \cdots j_1} \times \\ T_{r-1, j_1} \cdots T_{r-1, j_k} \frac{T_{r-1, r}}{\alpha_r} \cdot \frac{(1 + \alpha_r) C_{r-1}^{s-k-1} A^{k+1}}{r}$$

In view of the expression (44) of  $E_{s-r}^{(r-1)}$  we get (36) and (37). Here we need an estimate for  $\Theta(r, s, k)$ , that we defer to appendix A.3. Putting  $m = 2$  in (55) we get

$$(38) \quad \Theta(r, s - r, k) < 2^k r^{k-1} \left(1 + \frac{1}{r}\right)^k \binom{\frac{s}{r} - 2}{k-1}.$$

Furthermore, we isolate the contribution of the quantities  $T_{r-1, \cdot}$  that control the small divisors. With an appropriate use of property (ii) of lemma 8 we have

$$(39) \quad T_{r-1, j_1} \cdots T_{r-1, j_k} \frac{T_{r-1, r}}{\alpha_r} = \frac{1}{r^2} \left(\frac{\alpha_r}{r^2}\right)^{k-1} \frac{r^2 T_{r-1, j_1}}{\alpha_r} \cdots \frac{r^2 T_{r-1, j_{k-1}}}{\alpha_r} T_{r-1, j_k} \frac{r^2 T_{r-1, r}}{\alpha_r} \\ \leq \frac{1}{r^2} \left(\frac{\alpha_r}{r^2}\right)^{k-1} T_{r, s}.$$

Thus, using also  $\alpha_r \leq 1$  which is true in view of the definition of  $\alpha_r$ , we replace (36) with

$$\|E_{s-r}^{(r-1)} \mathbf{R}X_r\| \leq T_{r, s} C_{r-1}^{s-1} \frac{(1 + \alpha_r) A}{\alpha_r} \sum_{k=1}^{\lfloor s/r \rfloor - 1} \Theta(r, s - r, k) \left(\frac{\alpha_r A}{r^2 C_{r-1}}\right)^k \\ \leq T_{r, s} C_{r-1}^{s-1} A \cdot \frac{4 \left(1 + \frac{1}{r}\right) A}{r^2 C_{r-1}} \sum_{l=0}^{\lfloor s/r \rfloor - 2} \binom{\frac{s}{r} - 2}{l} \left(\frac{2 \left(1 + \frac{1}{r}\right) A}{r C_{r-1}}\right)^l \\ < T_{r, s} C_{r-1}^{s-1} A \cdot \frac{4 \left(1 + \frac{1}{r}\right) A}{r^2 C_{r-1}} \left(1 + \frac{2 \left(1 + \frac{1}{r}\right) A}{r C_{r-1}}\right)^{\frac{s}{r} - 2}.$$

For  $r = 1$  the latter formula readily gives (34). For  $r > 1$  we use  $1 + \frac{1}{r} < 2$  and  $\frac{8A}{C_{r-1}} < 1$ , which follows from the choice of the sequence  $C_r$  in the statement of the lemma, so that (35) is easily recovered.

We come now to the estimate of  $\|V_s^{(r)}\|$ . Here it is convenient to separate the case  $r = 1$ . Putting (34) and the induction hypothesis (31) in (22) we get

$$\begin{aligned} \|V_s^{(1)}\| &< T_{0,s} \frac{C_0^{s-1} A}{s} + T_{1,s} \frac{C_0^{s-1} A}{s} \cdot \frac{8A}{C_0} \left(1 + \frac{4A}{C_0}\right)^{s-2} \\ &< T_{1,s} \frac{C_0^{s-1} A}{s} \left[1 + \frac{8A}{C_0} \left(1 + \frac{4A}{C_0}\right)^{s-2}\right] \\ &< T_{1,s} \frac{C_0^{s-1} A}{s} \left(1 + \frac{8A}{C_0}\right)^{s-1}. \end{aligned}$$

Thus we may write

$$(40) \quad \|V_s^{(1)}\| < T_{1,s} \frac{\hat{C}_1^{s-1} A}{s}, \quad \hat{C}_1 = C_0 + 8A.$$

For  $r > 1$  we put (35) in (22), and we get

$$\begin{aligned} \|V_s^{(r)}\| &< T_{r-1,s} \frac{C_{r-1}^{s-1} A}{s} + T_{r,s} \frac{C_{r-1}^{s-1} A}{rs} \left(1 + \frac{1}{r}\right)^{\frac{s}{r}-2} \\ &< T_{r,s} \frac{C_{r-1}^{s-1} A}{s} \left[1 + \frac{1}{r} \left(1 + \frac{1}{r}\right)^{\frac{s}{r}-2}\right] < T_{r,s} \frac{C_{r-1}^{s-1} A}{s} \left(1 + \frac{1}{r}\right)^{\frac{s-1}{r}} \end{aligned}$$

Thus we conclude

$$(41) \quad \|V_s^{(r)}\| < T_{r,s} \frac{\hat{C}_r^{s-1} A}{s}, \quad \hat{C}_r = \left(1 + \frac{1}{r}\right)^{1/r} C_{r-1}.$$

Now we look for an estimate of  $\|W_s^{(r)}\|$  as given by (23). Recall that we have  $s > r$ , because  $W_r^{(r)} = 0$  by construction. We use (40) and (41) together with lemma 4, and get

$$\begin{aligned} \|W_s^{(r)}\| &\leq T_{r,s} \frac{\hat{C}_r^{s-1} A}{s} + \frac{1}{s} \sum_{k=1}^{\lfloor s/r \rfloor - 1} \frac{(s+2)(s-r+2) \cdots (s-kr+r+2)}{k!} \times \\ &\quad \left(\frac{r^2 T_{r-1,r}}{\alpha_r}\right)^k T_{r,s-kr} \times \left(\frac{C_{r-1}^{r-1} A}{r^2}\right)^k \hat{C}_r^{s-kr-1} A. \end{aligned}$$

Here we use again the statement (ii) of lemma 8. By the trivial inequality

$$s - jr + 2 \leq 4r \left(\frac{s}{r} - 1 - j\right) \quad \text{for } 0 \leq j < \lfloor s/r \rfloor - 1$$

and remarking that  $\hat{C}_r > 4A$  the latter estimate yields

$$\begin{aligned} \|W_s^{(r)}\| &\leq T_{r,s} \frac{\hat{C}_r^{s-1} A}{s} \left[ 1 + \sum_{k=1}^{\lfloor s/r \rfloor - 1} \binom{\frac{s}{r} - 1}{k} \left( \frac{4A}{r^2 \hat{C}_r} \right)^k \right] \\ &< T_{r,s} \frac{\hat{C}_r^{s-1} A}{s} \sum_{k=0}^{\lfloor s/r \rfloor - 1} \binom{\frac{s}{r} - 1}{k} \left( \frac{1}{r^2} \right)^k \\ &< T_{r,s} \frac{\hat{C}_r^{s-1} A}{s} \left( 1 + \frac{1}{r^2} \right)^{\frac{s-1}{r}}. \end{aligned}$$

We conclude

$$(42) \quad \|W_s^{(r)}\| < T_{r,s} \frac{C_r^{s-1} A}{s}, \quad C_r = \left( 1 + \frac{1}{r^2} \right)^{1/r} \hat{C}_r.$$

In view of (40) and (41) this proves the claim of the lemma. Q.E.D.

## 4. Proof of the main theorem

Having established the estimate of the iteration lemma 10 on the sequence of generating vector fields it is now a standard matter to complete the proof of theorem 1. Hence this section will be less detailed with respect to the previous ones.

In view of the iteration lemma we are given an infinite sequence  $\{X_r\}_{r \geq 1}$  of generating vector fields with  $X_r$  homogeneous polynomial of degree  $r + 1$  satisfying

$$\|X_r\| \leq T_{r-1,r} \frac{C_{r-1}^{r-1} A}{r \alpha_r}.$$

By lemma 9 we have  $\frac{T_{r-1,r}}{\alpha_r} \leq \gamma^r e^{r\Gamma}$  with  $\Gamma$  as in condition  $\tau$  and  $\gamma$  a constant independent of  $\lambda$ . Moreover, still by lemma 10, we have  $C_r \leq C_\infty$ , a positive constant independent of  $\lambda$ . Thus we have

$$(43) \quad \|X_r\| \leq \frac{\eta^r e^{r\Gamma}}{r} K$$

with positive constants  $\eta$  and  $K$  independent of  $\lambda$ .

We refer now to the analytic setting that we recall in appendix A.2. Every vector field  $X_r$  generates a near the identity transformation  $y = \exp(L_{X_r})x$  which transforms the generating sequence  $W^{(r-1)}$  into  $W^{(r)}$  according to the algorithm of section 2.2. Thus the near the identity transformation to normal form is generated by the limit  $S_X$  of the sequence of operators  $S_X^{(r)} = \exp(L_{X_r}) \circ \dots \circ \exp(L_{X_1})$ . We apply proposition 2 of appendix A.2. In view of (43), in a polydisk  $\Delta_\varrho$  of radius  $\varrho$  centered at the origin of  $\mathbb{C}^n$  we have

$$|X_r|_\varrho \leq \|X_r\| \varrho^{r+1} \leq \frac{\eta^r e^{r\Gamma}}{r} K \varrho^{r+1},$$

where  $|X_r|_\varrho$  is the supremum norm. Thus condition (52) of proposition 2 reads

$$\varrho \sum_{r \geq 1} \frac{\eta^r e^{r\Gamma}}{r} K \varrho^r < \frac{\varrho}{4eK} ,$$

which is true if we take, e.g.,  $\varrho < \bar{\varrho} = \frac{3}{2} B^{-1} e^{-\Gamma}$  with a constant  $B$  independent of  $\lambda$ . Thus proposition 2 applies with, e.g.,  $\delta = \bar{\varrho}/3$ , and we conclude that the near the identity transformation that gives the map the wanted linear normal form is analytic at least in a polydisk of radius  $B^{-1} e^{-\Gamma}$ , as claimed. This concludes the proof of the main theorem.

## Appendix

### A. Technicalities

This appendix is mainly devoted to recalling some definitions and properties related to Lie series and Lie transforms that are used in the paper. The estimate of the sequence (37) is also included.

#### A.1 Lie series and Lie transforms

We recall here some formal properties of the Lie series and Lie transforms as defined by (8) and (9). The Lie derivative with respect to the vector field  $X$  is defined as

$$L_X \cdot = \left. \frac{d}{dt} (\phi_X^t \cdot) \right|_{t=0}$$

where  $\phi_X^t$  is the flow generated by  $X$ . We also recall that the explicit form of the Lie derivatives for a function  $f(x)$  and a vector field  $v(x)$  are, respectively,

$$L_X f = \sum_{j=1}^n X_j \frac{\partial f}{\partial x_j} \quad \text{and} \quad (L_X v)_j = \sum_{l=1}^n \left( X_l \frac{\partial v_j}{\partial x_l} - v_l \frac{\partial X_j}{\partial x_l} \right) ,$$

where the l.h.s. is the  $j$ -th component of the vector field  $L_X v$ .

A non-recursive formula for the operator  $E_s^X$  defined by (10) is given by

**Lemma 11:** *Let  $X = \{X_l\}_{l>1}$  be a sequence of vector fields and  $T_X = \sum_{s \geq 0} E_s^X$  as in (10). Then for  $s > 0$  we have*

$$(44) \quad E_s^X = \sum_{k=1}^s \sum_{\substack{j_1 + \dots + j_k = s \\ j_1, \dots, j_k > 0}} \frac{j_k \cdots j_1}{(j_1 + \dots + j_k) \cdot (j_1 + \dots + j_{k-1}) \cdots j_1} L_{X_{j_k}} \circ \dots \circ L_{X_{j_1}} .$$

If moreover  $X_1 = \dots = X_{r-1} = 0$  then we get the same formula with the limitation  $j_1, \dots, j_k \geq r$  on the indices  $j$ .

**Proof.** The formula may be written down directly by combinatorial considerations. A proof by induction goes as follows. We write the denominator in the equivalent form



$s(s - j_k) \cdots (s - j_k - \dots - j_2)$ . For  $s = 1$  both the formula and the recursive definition (10) give  $E_1^X = L_{X_1}$ . Assuming it is true up to  $s - 1$ , for  $l = 1, \dots, s - 1$  write the corresponding term in the r.h.s. of (10) as

$$\frac{l}{s} L_{X_l} E_{s-l}^{(X)} = \sum_{k=1}^{s-l} \sum_{\substack{j_1 + \dots + j_k = s-l \\ j_1, \dots, j_k > 0}} \frac{l}{s} L_{X_l} \frac{j_k \cdots j_1}{(s-l)(s-l-j_k) \cdots (s-l-j_k - \dots - j_2)} L_{X_{j_k}} \circ \dots \circ L_{X_{j_1}}$$

while for  $l = s$  we have the sole term  $L_{X_s}$ . Renaming the index  $l$  as  $j_{k+1}$  the corresponding term  $\frac{l}{s} L_{X_l}$  is included in the sum over the indexes  $j_1, \dots, j_{k+1}$  with the condition  $j_1 + \dots + j_{k+1} = s$ , and writing  $k$  in place of  $k+1$  the limits of the first sum run from 2 to  $s$ , thus changing the sum into  $\sum_{k=2}^s \sum_{j_1, \dots, j_k} \cdots$  so that only the term  $k = 1$  is missing. Adding it means that the second sum contains the sole term  $j = s$ , corresponding to  $L_{X_s}$ , which is actually the missing term in order to recover (10). *Q.E.D.*

The Lie series and Lie transform are linear operators acting on the space of holomorphic functions and of holomorphic vector fields. They preserve products between functions and commutators between vector fields, i.e., if  $f, g$  are functions and  $v, w$  are vector fields then one has

$$(45) \quad T_X(fg) = T_X f \cdot T_X g, \quad T_X\{v, w\} = \{T_X v, T_X w\}.$$

Here, replacing  $T_X$  with  $\exp(L_X)$  gives the corresponding property for Lie series. Moreover both operators are invertible. The inverse of  $\exp(L_X)$  is  $\exp(L_{-X})$ , which is a natural fact if one recalls the origin of Lie series as a solution of an autonomous system of differential equations.

We come now to a remarkable property which justifies the usefulness of Lie methods in perturbation theory. We adopt the name *exchange theorem* introduced by Gröbner. Let  $f$  be a function and  $v$  be a vector field. Consider the transformation  $y = T_X x$ , i.e., in coordinates,

$$y_j = x_j + X_{1,j}(x) + \left[ \frac{1}{2} L_{X_1} X_{1,j}(x) + X_{2,j}(x) \right] + \dots, \quad j = 1, \dots, n.$$

and denote by  $J$  its differential, namely, in coordinates, the Jacobian matrix with elements  $J_{j,k} = \frac{\partial y_j}{\partial x_k}$ . Then one has

$$(46) \quad f(y)|_{y=T_X x} = (T_X f)(x), \quad J^{-1}v(y)|_{y=T_X x} = (T_X v)(x).$$

The same statement holds true for a transformation generated by a Lie series.

The main difference between Lie series and Lie transform is that a near the identity transformation of the form (2) can be represented via a Lie transform with a suitable sequence of vector fields. The same does not hold true for a Lie series, but the same result may be achieved by using the composition of Lie series. Again, see [11] for details.

## A.2 Analyticity of the compositions of Lie series

We include here some quantitative estimates which lead to the convergence of a near the identity coordinate transformation defined by composition of Lie series. The results here are basically an extension of Chapter 1 of Gröbner's book [13]. The reader may also find more details in [7] where the Hamiltonian case is dealt with.

We consider a polydisk of radius  $\varrho$  centered at the origin

$$\Delta_\varrho = \{x \in \mathbb{C}^n : |x| \leq \varrho\} .$$

For a function  $f$  and a vector field  $X$  it is convenient to consider the usual supremum norm

$$|f|_\varrho = \sup_{x \in \Delta_\varrho} |f(x)| \quad \text{and} \quad |X|_\varrho = \sum_j |X_j|_\varrho .$$

The following lemma states the Cauchy inequality for Lie derivatives in  $\Delta_\varrho$ .

**Lemma 12:** *For  $0 \leq \delta' < \delta < \varrho$  the following inequalities apply to Lie derivatives:*

$$|L_X f|_{\varrho-\delta} \leq \frac{1}{\delta - \delta'} |X|_\varrho |f|_{\varrho-\delta'} ,$$

and, for  $s \geq 1$ ,

$$|L_X^s f|_{\varrho-\delta} \leq \frac{s!}{e} \left( \frac{e|X|_\varrho}{\delta} \right)^s |f|_\varrho .$$

The proof is a straightforward adaptation of Cauchy estimates for derivatives of analytic functions. The reader will be able to reproduce it himself, perhaps with the help of the proof for the Hamiltonian case in [7].

The following proposition essentially restates Cauchy's existence theorem of solution of differential equations in the analytic case.

**Proposition 1:** *Let  $X$  be an analytic vector field on the interior of  $\Delta_\varrho$  and bounded on  $\Delta_\varrho$ . Then, for every positive  $\delta \leq \varrho/2$  the following statement holds true: if  $|X|_\varrho < \delta(e-1)/e^2$  then the map*

$$(47) \quad \varphi(x) = \exp(L_X)x$$

is bianalytic and satisfies

$$(48) \quad |\exp(L_X)x_j - x_j|_{\varrho-\delta} \leq \frac{\delta}{e^2} ,$$

so that we have

$$\Delta_{\varrho-2\delta} \subset \varphi(\Delta_{\varrho-\delta}) \subset \Delta_\varrho , \quad \Delta_{\varrho-2\delta} \subset \varphi^{-1}(\Delta_{\varrho-\delta}) \subset \Delta_\varrho .$$

**Proof.** Using lemma 12, compute

$$\begin{aligned} |\exp(L_X)x_j - x_j|_{\varrho-\delta} &= \sum_{s \geq 1} \left| \frac{L_X^{s-1}}{s!} X_j \right|_{\varrho-\delta} \\ &\leq \sum_{s \geq 1} \frac{1}{e s} \left( \frac{e|X|_\varrho}{\delta} \right)^{s-1} |X|_\varrho \leq -\frac{\delta}{e^2} \ln \left( 1 - \frac{e|X|_\varrho}{\delta} \right) \leq \frac{\delta}{e^2} , \end{aligned}$$

namely (48). The same estimates apply to the inverse  $\exp(L_{-X})$ .

*Q.E.D.*

Let now  $\{X_1, X_2, \dots\}$  be a sequence of analytic vector fields. Consider the sequence of transformations  $\{S_X^{(0)}, S_X^{(1)}, S_X^{(2)}, \dots\}$  recursively defined as

$$(49) \quad S_X^{(0)} = 1, \quad S_X^{(r)} = \exp(L_{X_r}) \circ S_X^{(r-1)}.$$

with inverse

$$(50) \quad \tilde{S}_X^{(0)} = 1, \quad \tilde{S}_X^{(r)} = \tilde{S}_X^{(r-1)} \circ \exp(L_{-X_r}).$$

We may well consider in formal sense the composition of Lie series as the limit

$$(51) \quad S_X = \dots \circ \exp(L_{X_r}) \circ \dots \circ \exp(L_{X_2}) \circ \exp(L_{X_1}),$$

with a similar formal limit for the sequence  $\tilde{S}_X^{(r)}$  giving the inverse  $S_X^{-1}$ .

**Proposition 2:** *Let the sequence  $X = \{X_r\}_{r \geq 1}$  of vector fields be analytic in  $\Delta_\varrho$ . Then for every positive  $\delta < \varrho/2$  the following statement holds true: if the convergence condition*

$$(52) \quad \sum_{r \geq 1} |X_r|_\varrho < \frac{\varrho}{4e}$$

*is satisfied then the sequence  $\{S_X^{(0)}, S_X^{(1)}, S_X^{(2)}, \dots\}$  of maps defined by (49) is bianalytic and converges to a bianalytic map  $S_X$  satisfying*

$$(53) \quad \Delta_{\varrho-2\delta} \subset S_X(\Delta_{\varrho-\delta}) \subset \Delta_\varrho, \quad \Delta_{\varrho-2\delta} \subset S_X^{-1}(\Delta_{\varrho-\delta}) \subset \Delta_\varrho.$$

**Proof.** Consider the sequence of positive numbers

$$\delta_s = \frac{|X_s|_\varrho}{\sum_{r \geq 1} |X_r|_\varrho} \delta, \quad s \geq 1$$

so that  $\sum_{s \geq 1} \delta_s = \delta$ . By definition we have  $|X_s|_\varrho < \frac{\delta_s}{2e}$ , so that proposition 1 applies to every vector field  $X_s$ . Consider also the two sequences

$$\varrho_0^- = \varrho_0^+ = \varrho - \delta, \quad \varrho_s^- = \varrho_{s-1}^- - \delta_s, \quad \varrho_s^+ = \varrho_{s-1}^+ + \delta_s, \quad s \geq 1,$$

so that we have  $\varrho_s^- \xrightarrow{s \rightarrow +\infty} \varrho - 2\delta$  and  $\varrho_s^+ \xrightarrow{s \rightarrow +\infty} \varrho$ . Applying proposition 1 we see that the map  $S^{(1)}$  is analytic and satisfies  $\Delta_{\varrho_1^-} \subset S^{(1)}\Delta_{\varrho_0} \subset \Delta_{\varrho_1^+}$ , and that its inverse is analytic and satisfies  $\Delta_{\varrho_1^-} \subset S^{(1)}\Delta_{\varrho_0} \subset \Delta_{\varrho_1^+}$ . Proceeding by induction it is straightforward to check that the same claim holds true for  $S^{(r)}$  and  $\tilde{S}^{(r)}$  for every  $r > 1$ , with  $\varrho_1^-$  and  $\varrho_1^+$  replaced by  $\varrho_r^-$  and  $\varrho_r^+$ , respectively. Moreover the inequalities  $|S^{(r)}x - S^{(r-1)}x| < \delta_r$  and  $|\tilde{S}^{(r)}x - \tilde{S}^{(r-1)}x| < \delta_r$  hold true. Thus the sequences  $S^{(r)}$  and  $\tilde{S}^{(r)}$  of maps is uniformly convergent in any compact subset of  $\Delta_{\varrho_0} = \Delta_{\varrho-\delta}$ , and by Weierstrass theorem the limits  $S_x$  and  $S_X^{-1}$  do exist and are analytic maps satisfying (53). *Q.E.D.*

### A.3 Estimate of the sequence (37)

We introduce a parameter  $m \geq 0$  which is arbitrary but fixed and rewrite the sequence (37) as

$$(54) \quad \Theta(r, s, k) = \sum_{\substack{j_1 + \dots + j_k = s \\ j_1, \dots, j_k \geq r}} \frac{(j_1 + \dots + j_k + r + m)(j_1 + \dots + j_{k-1} + r + m) \cdots (j_1 + r + m)}{(j_1 + \dots + j_k)(j_1 + \dots + j_{k-1}) \cdots j_1}$$

with  $s \geq kr$  and  $1 \leq k \leq \lfloor s/r \rfloor$ . With a straightforward calculation we get the uniform estimate

$$\begin{aligned} & \frac{(j_1 + \dots + j_k + r + m)(j_1 + \dots + j_{k-1} + r + m) \cdots (j_1 + r + m)}{(j_1 + \dots + j_k)(j_1 + \dots + j_{k-1}) \cdots j_1} \\ &= \left(1 + \frac{r + m}{j_1 + \dots + j_k}\right) \cdot \left(1 + \frac{r + m}{j_1 + \dots + j_{k-1}}\right) \cdots \left(1 + \frac{r + m}{j_1}\right) \\ &< \left(2 + \frac{m}{r}\right)^k, \end{aligned}$$

which holds true because  $j_1, \dots, j_k \geq r$ . Thus we get

$$\Theta(r, s, k) = \sum_{\substack{j_1 + \dots + j_k = s \\ j_1, \dots, j_k \geq r}} \left(2 + \frac{m}{r}\right)^k.$$

Writing  $j_1 = r + q_1, \dots, j_k = r + q_k$  the number of terms in the sum is equal to the number of non negative integer vectors  $q = (q_1, \dots, q_n)$  with  $|q| = s - kr$ , which is known to be

$$\binom{s - kr + k - 1}{k - 1} \leq r^{k-1} \binom{\frac{s}{r} - 1}{k - 1}.$$

We conclude

$$(55) \quad \Theta(r, s, k) < r^{k-1} \left(2 + \frac{m}{r}\right)^k \binom{\frac{s}{r} - 1}{k - 1}.$$

### Acknowledgement

This research is partially supported by PRIN-MIUR 2010JJ4KPA “Teorie geometriche e analitiche dei sistemi Hamiltoniani in dimensioni finite e infinite”. M. S. has also been partially supported by an FSR Incoming Post-doctoral Fellowship of the Académie universitaire Louvain, co-funded by the Marie Curie Actions of the European Commission.

### References

- [1] V.I. Arnold: *Geometrical Methods in the Theory of Ordinary Differential Equations*, Springer Verlag, (1980)
- [2] Buff, X. and Chéritat, A.: *The Brjuno function continuously estimates the size of quadratic Siegel disks*, Ann. of Math. **164**, 265–312 (2006).
- [3] Carletti, T. and Marmi, S.: *Linearization of analytic and non-analytic germs of diffeomorphisms of  $(\mathbb{C}, 0)$* , Bull. Soc. Math. France, **128**, 69–85 (2000).

- [4] Giorgilli, A. and Locatelli, U.: *On classical series expansions for quasi-periodic motions*, MPEJ **3** N. 5 (1997).
- [5] A. Giorgilli and U. Locatelli: *A classical self-contained proof of Kolmogorov's theorem on invariant tori*, in *Hamiltonian systems with three or more degrees of freedom*, Carles Simó ed., NATO ASI series C, Vol. 533, Kluwer Academic Publishers, Dordrecht–Boston–London (1999).
- [6] Giorgilli, A.: *Unstable equilibria of Hamiltonian systems*, Disc. and Cont. Dynamical Systems, Vol. 7, N. 4, 855–871 (2001).
- [7] A. Giorgilli: *Notes on exponential stability of Hamiltonian systems*, in *Dynamical Systems, Part I: Hamiltonian systems and Celestial Mechanics* Pubblicazioni del Centro di Ricerca Matematica Ennio De Giorgi, Pisa, 87–198 (2003).
- [8] A. Giorgilli and D. Muraro: *Exponentially stable manifolds in the neighbourhood of elliptic equilibria*, Boll. Unione Mat. Ital. Sez. B **9**, 1–20 (2006).
- [9] A. Giorgilli and S. Marmi: *Convergence radius in the Poincaré-Siegel problem*, DCDS Series S, **3**, 601–621 (2010).
- [10] A. Giorgilli: *On a theorem of Lyapounov*, Rendiconti dell'Istituto Lombardo Accademia di Scienze e Lettere, Classe di Scienze Matematiche e Naturali, to appear (see also arXiv:1303.7322).
- [11] A. Giorgilli: *On the representation of maps by Lie transforms*, Rendiconti dell'Istituto Lombardo Accademia di Scienze e Lettere, Classe di Scienze Matematiche e Naturali, to appear (see also arXiv:1211.5674).
- [12] A. Giorgilli, U. Locatelli and M. Sansottera: *On the convergence of an algorithm constructing the normal form for lower dimensional elliptic tori in planetary systems*, (arXiv:1401.6529).
- [13] W. Gröbner: *Die Lie-Reihen und Ihre Anwendungen*, VEB Deutscher Verlag der Wissenschaften, Mathematische Monographien, **3** (1960). Italian translation: *Serie di Lie e loro applicazioni*, Ed. Cremonese, Roma (1973).
- [14] E. Schröder: *Über iterierte Functionen*, Math. Ann. **3**, 296–322 (1871).
- [15] Siegel, C.L.: *Iterations of analytic functions*, Annals of Math. **43**, 607–612 (1942).
- [16] Yoccoz, J.-C.: *Théorème de Siegel, nombres de Bruno et polynômes quadratiques* Astérisque **231** 3–88 (1995).
- [17] Yoccoz, J.-C.: *Analytic linearisation of circle diffeomorphisms* in “Dynamical Systems and Small Divisors”, Lecture Notes in Mathematics **1784** 125–173 (2002).