

# IMPROVED ESTIMATES FOR THE CONVERGENCE RADIUS IN THE POINCARÉ–SIEGEL PROBLEM

*ANTONIO GIORGILLI*

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

*STEFANO MARMI*

*Scuola Normale Superiore  
Piazza dei Cavalieri 7, 56126 Pisa, Italy.*

**Abstract.** We reconsider the Poincaré–Siegel center problem, namely the problem of conjugating an analytic system of differential equations in the neighbourhood of an equilibrium to its linear part. 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  $r$  of the conjugating transformation satisfies  $\log r(\lambda) \geq -C\mathcal{B} + C'$  with  $C = 1$  and a constant  $C'$  not depending on  $\lambda$ . This improves the previous results for  $n > 1$ , where the known proofs gives  $C = 2$ . 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 old standing problem of convergence of perturbation series involving small divisors with the aim of producing general and possibly optimal estimates for the radius of convergence.

Small divisor problems arise naturally when nonlinear quasiperiodic dynamical systems are considered. A classical example is given by the iteration of analytic maps having the origin as a fixed point, when the eigenvalues of the linear part of the map lie on the unitary circle. When an irrational rotation in  $\mathbb{C}^n$  is analytically perturbed, it is natural to ask whether or not there exists a neighborhood of the fixed point where the dynamics looks like the unperturbed case. This question can be made more precise looking for a local holomorphic coordinate change such that in the new coordinates the perturbed vector field is expressed as an ordinary rotation, i.e. it becomes linear. The formal solution of this problem in dimension one is given by Schröder series<sup>[25]</sup>, which represent the simplest exemple of perturbation series involving small divisors. The analyticity of Schröder series has been proved by Siegel in 1942. He used the majorant series method and a delicate number-theoretical lemma, often called Siegel’s lemma, putting particular emphasys on the relevance of Diophantine approximations<sup>[26]</sup>. This was actually the first proof of convergence of a meaningful small divisor series.

Ten years later Siegel himself succeeded in applying the same ideas to the so called Poincaré center problem, namely the problem of conjugating an analytic system of differential equations in the neighbourhood of an equilibrium to its linear part<sup>[27]</sup>. The problem was first solved by Poincaré by assuming that there exists in the complex plane a straight line separating the origin from the eigenvalues of the linear part of the system<sup>[23]</sup>. Siegel's proof deals with the case in which Poincaré's condition is not fulfilled, which makes the problem of small divisors a major one. In the present paper we shall reconsider precisely this problem.

In 1954 Kolmogorov announced his celebrated theorem on the persistence under small perturbations of quasiperiodic motions of integrable Hamiltonian systems<sup>[14]</sup>. The diophantine irrationality condition on the frequencies plays a central role in his proof. Moreover he introduced an iteration scheme, in his own words "similar to Newton's method", assuring a fast convergence. This method, often called superconvergent or quadratic, has been used by some authors as a replacement for the majorant's method in order to give new proofs of the Siegel's theorems mentioned above: see, e.g., [28]. The subsequent works of Moser<sup>[20][21]</sup> and Arnold<sup>[1][2]</sup> on area preserving mappings and Hamiltonian systems marked the beginning of the so called KAM theory.

As a matter of fact our understanding of the convergence phenomena concerned with small divisors is still far from being complete<sup>[18]</sup>. For example, restricting the attention to near to integrable Hamiltonian systems, or to perturbations of translations on tori, we still do not know how to characterize *exactly* the set of rotation vectors  $\omega$  for which an invariant torus carrying quasiperiodic motions of frequency  $\omega$  persists under a (sufficiently small) analytic perturbation.

The previous question has been answered in only one case, namely the problem of linearization of analytic one-frequency systems. Exploiting the geometric renormalization approach some spectacular results have been obtained in the last 20 years: see [29], [22], [30], and [6]. In this case the optimal set of rotation numbers for which an analytic linearization exists is known, being given by the set of *Bruno numbers*. The same set plays an analogue role for some area-preserving maps [16], [8], including the standard family<sup>[17][3][4]</sup>. However such result do not generalize to higher dimension.

Bruno numbers were first introduced by A.D. Bruno<sup>[5]</sup> (see also [24]). More generally, Bruno gave a proof of Siegel's theorem for Poincaré's center problem assuming a nonresonance condition weaker than the diophantine one: we shall discuss the condition of Bruno in section 5.1, formula (26).

Bruno's proof gives an estimate from below of the radius of convergence  $r$  of the linearizing transformation, getting  $\log r(\Lambda) \geq -CB + C'$  where  $C = 2$  and  $C'$  is a universal constant (i.e. independent of the eigenvalues of the linear part  $\Lambda$ ). In [7] it has been showed that when  $n = 1$  one can adapt the Siegel-Bruno majorant series method and obtain  $C = 1$ . The same lower bound had been proved by geometric renormalization in [29] where the question of whether the majorant series method were able to obtain the same bound was explicitly asked.

In this paper we show that the same bound with the optimal constant  $C = 1$  can be proved in any dimension  $n$  by implementing a majorant series method based on generation of coordinate changes using Lie series. We stress that the interest of the

method is not limited to the case of Poincaré center problem. E.g., for an application to Kolmogorov’s case see [10] or [11].

Let us come to a formal statement of our result. We consider a system of differential equations

$$(1) \quad \dot{x} = \Lambda x + v(x)$$

where  $x \in \mathcal{U} \subset \mathbb{C}^n$  with  $\mathcal{U}$  an open neighbourhood of the origin of  $\mathbb{C}^n$ , and  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$  is a diagonal matrix. We assume that  $v : \mathcal{U} \rightarrow \mathbb{C}^n$  is a vector field analytic on the polydisk  $|x_i| < b^{-1}$ ,  $i = 1, \dots, n$ , where  $b$  is some positive. Moreover we will assume that  $v$  satisfies

$$(2) \quad v(0) = 0, \quad dv(0) = 0.$$

For integer  $r \geq 0$  let us define

$$(3) \quad \beta_0 = 1, \quad \beta_r = \min_{j=1, \dots, n} \min_{k \in \mathbb{Z}_+^n, |k|=r+1} |\langle k, \lambda \rangle - \lambda_j|.$$

Furthermore let us define the non-increasing sequence of real numbers

$$(4) \quad \alpha_r = \min_{0 \leq s \leq r} \beta_s \quad \text{for } r \geq 0.$$

We shall say that the eigenvalues  $\lambda_1, \dots, \lambda_n$  are non resonant in case  $\beta_r > 0$  for all  $r$ . This obviously implies that  $\alpha_r > 0$  for all  $r \geq 0$ . Actually we shall require the stronger

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

$$(5) \quad - \sum_{r \geq 1} \frac{\ln \alpha_r}{r(r+1)} = \Gamma < \infty.$$

**Theorem 1:** *Let  $v$  be as above and assume that the eigenvalues of  $\Lambda$  are nonresonant and satisfy condition  $\tau$ . Then there exists a near to identity coordinate transformation  $x = y + \psi(y)$ , with  $\psi$  analytic on the polydisk of radius  $b^{-1}A^{-1}e^{-\Gamma}$ , where  $A > 0$  is a universal constant, which transforms equation (1) into  $\dot{y} = \Lambda y$ .*

The paper is organized as follows. In Section 2 we briefly recall the use of Lie series for implementing coordinate transformations. Since this is now a quite standard matter, most technical details are deferred to Appendix A. Section 3 will be devoted to understanding how the small divisors accumulate through the normalization steps introduced in Section 2. Our characterization of the accumulation of small divisors constitutes the most original part of our work. We actually show that the approach of Siegel and Bruno can be improved so as to obtain an estimate on the contribution of small divisors which holds in arbitrary dimension and is known to be optimal in the one frequency case. The linearization result will be proved in Section 4, whereas Section 5 will be devoted to the comparison between condition  $\tau$  and the Bruno function. The Appendixes will be devoted to some (quite standard) technical proofs and to a brief summary of elementary facts on continued fractions.

## 2. Technical tools and normal form

In this section we briefly recall the use of Lie series in order to give the system (1) a normal form. The algorithm is first developed at a formal level. Then we recall the quantitative estimates that represent the technical tools for the discussion of the convergence.

### 2.1 Formal algorithm

We shall consider transformations defined through the time–one flow generated by a vector field  $X = (X_1, \dots, X_n)$ , where  $X_1, \dots, X_n$  are homogeneous polynomials of degree  $r \geq 2$ . Then an analytic vector field  $v(x)$  is transformed as

$$(6) \quad v' = \exp(L_X v) = v + L_X v + L_X^2 v + \dots ,$$

where  $L_X$  is the Lie derivative along the flow  $\phi_X^t$  generated by  $X$ , namely

$$L_X v = \frac{d}{dt} (v \circ \phi_X^t)(x) \Big|_{t=0} .$$

Actually we shall use the expression of the Lie derivative in coordinates, namely

$$(7) \quad v'_j = \sum_{k=1}^n \left( X_k \frac{\partial v_j}{\partial x_k} - v_k \frac{\partial X_j}{\partial v_k} \right) ,$$

i.e., the commutator between the vector fields  $X$  and  $v$ .

Let us now come to the formal normalization algorithm. We expand the vector field in equation (1) in power series around the origin as

$$v(x) = v_1^{(0)}(x) + v_2^{(0)}(x) + v_3^{(0)}(x) + \dots$$

where  $v_s^{(0)}(x)$  is homogeneous polynomial of degree  $s + 1$ , the upper index 0 meaning that no normalization steps have been performed till now. After  $r - 1$  normalization steps the system will be given the Poincaré's normal form up to order  $r$

$$\dot{x} = \Lambda x + v_r^{(r-1)} + v_{r+1}^{(r-1)} + \dots ,$$

meaning that the power series expansion of the still non normalized part of the vector field starts with terms of degree  $r + 1$ . The upper index indicates that  $r - 1$  normalization steps have been performed.

The normalization step from order  $r - 1$  to order  $r$  is worked out as follows. We consider the generating vector field  $X_r$  which solves the homological equation

$$(8) \quad D_\lambda X_r = v_r^{(r-1)} ,$$

where  $D_\lambda = L_{\Lambda x}$  is the commutator with the linear vector field  $\Lambda x$ . In view of the non–resonance condition satisfied by the eigenvalues  $\lambda_1, \dots, \lambda_n$  the latter equation admits a unique solution which is a homogeneous polynomial of degree  $r + 1$ . By giving the

transformation by Lie series a more explicit expression the transformed vector field is constructed as

$$(9) \quad \begin{aligned} v_{sr+m}^{(r)} &= \sum_{p=0}^{s-1} \frac{1}{p!} L_{X_r}^p v_{(s-p)r+m}^{(r-1)} \quad \text{for } r \geq 2, s \geq 1 \text{ and } 1 \leq m < r; \\ v_{sr}^{(r)} &= \frac{s-1}{s!} L_{X_r}^{s-1} v_r^{(r-1)} + \sum_{p=0}^{s-2} \frac{1}{p!} L_{X_r}^p v_{(s-p)r}^{(r-1)} \quad \text{for } r \geq 1 \text{ and } s \geq 2. \end{aligned}$$

For  $r = 1$  only the latter expression is used, while we need also the first one at all the subsequent steps, due to the fact that the linear operator  $L_{X_r}$  maps polynomial vector fields of degree  $s + 1$  to polynomial vector fields of degree  $r + s + 1$ . Finding the expressions above requires some calculations that we defer to appendix A.1.

## 2.2 Quantitative estimates

We come now to introducing quantitative tools. Write a homogeneous polynomial of degree  $s$  in multiindex notation as  $f(x) = \sum_{|k|=s} f_k x^k$ , where  $k \in \mathbb{Z}_+^n$  is an integer vector with non-negative components and  $|k| = k_1 + \dots + k_n$  is the degree of the monomial  $x^k = x_1^{k_1} \cdot \dots \cdot x_n^{k_n}$ . Let  $R = (R_1, \dots, R_n) \in \mathbb{R}^n$  be a vector with positive components. We define the *polynomial norm* of  $f$  parameterized by  $R$  as

$$(10) \quad \|f\|_R = \sum_{|k|=r} |f_k| R^k,$$

For a vector field  $v = (v_1, \dots, v_n)$ , we define the norm as

$$(11) \quad \|v\|_R = \max_j \|v_j\|_R,$$

using the polynomial norm for the components.

The following technical lemma is a generalization of Cauchy estimates for the derivatives of an analytic function to the case of Lie derivatives.

**Lemma 1:** *Let  $X$  and  $v$  be homogeneous polynomial vector fields. Let  $0 \leq d' < d < 1$  and assume that  $\|X\|_R$  and  $\|v\|_{(1-d')R}$  are known. Let also  $Q = \min_j R_j$ . Then for every  $s \geq 1$  we have*

$$(12) \quad \|L_X^s v\|_{(1-d)R} \leq \frac{s!}{e} \left( \frac{2e\|X\|_R}{(d-d')Q} \right)^s \|v\|_{(1-d')R}.$$

The proof is deferred to appendix A.2.

A second lemma is concerned with the solution of the homological equation (8).

**Lemma 2:** *Let  $\psi(x)$  be a homogeneous polynomial vector field of degree  $r \geq 2$ , and let  $\lambda \in \mathbb{C}^n$  be a non resonant vector. Then the homological equation*

$$(13) \quad D_\lambda X = \psi$$

admits a unique solution  $X$  which is homogeneous polynomial of degree  $r$ , and moreover we have

$$(14) \quad \|X\|_R \leq \frac{1}{\beta_r} \|\psi\|_R ,$$

with  $\beta_r$  defined by (3).

**Proof.** Write the components of the vector field as

$$\psi_j = \sum_{|k|=r} \psi_{j,k} x^k , \quad X_j = \sum_k c_{j,k} x^k ,$$

where  $\psi_{j,k} \in \mathbb{C}$  are known coefficient and  $c_{j,k} \in \mathbb{C}$  are the coefficients to be determined. Recalling that  $D_\lambda = L_{\Lambda x}$  is the commutator with the linear vector field  $\Lambda x$  we readily calculate

$$\begin{aligned} (D_\lambda X)_j &= \sum_{l=1}^n \left( (\Lambda x)_l \frac{\partial X_j}{\partial x_l} - X_l \frac{\partial (\Lambda x)_j}{\partial x_l} \right) \\ &= \sum_k (\langle k, \lambda \rangle - \lambda_j) c_{j,k} x^k . \end{aligned}$$

This shows that  $D_\lambda$  is diagonal on the basis of monomials. Thus the unique vector field  $X$  satisfying (13) is determined as

$$X_j = \sum_{|k|=r} \frac{\psi_{j,k}}{\langle k, \lambda \rangle - \lambda_j} x^k ,$$

so that  $X$  is homogeneous polynomial, as claimed. From this the estimate (14) readily follows in view of the definition (3) of  $\beta_r$  and the definition of the norm of a homogeneous polynomial vector field. Q.E.D.

Using the lemmas above we may set up a scheme of recursive estimates for the normal form. To this end let us pick a positive  $d < 1$  and consider a sequence  $d_1, d_2, d_3, \dots$  of positive numbers satisfying  $\sum_{j \geq 1} d_j = d$ . Let us also set  $\delta_0 = 0$  and  $\delta_r = \sum_{l=1}^r d_l$ .

**Lemma 3:** For  $\|v_s^{(r)}\|_{(1-\delta_r)R}$  as defined by (9) the following recursive estimate holds true:

$$(15) \quad \|v_{sr+m}^{(r)}\|_{(1-\delta_r)R} \leq \frac{1}{e} \sum_{p=0}^{s-1} \left( \frac{2e \|v_r^{(r-1)}\|_{(1-\delta_{r-1})R}}{d_r \beta_r Q} \right)^p \|v_{(s-p)r+m}^{(r-1)}\|_{(1-\delta_{r-1})R}$$

for  $r \geq 1$ ,  $s \geq 1$  and  $0 \leq m < r$ , excluding  $(s = 1, m = 0)$ .

The proof is a straightforward translation of the recursive relations (9) into recursive estimates for the norm. This is easily done by using the estimates in lemmas 1 and 2.

### 3. Small divisors and selection rules

This section is devoted to the analysis of the accumulation of small divisors in the

recursive estimate of lemma 3. Since it contains the main novelties with respect to previously published proofs of Siegel’s theorem we include a detailed discussion.

The guiding remark is that the divisors to be taken into account are products  $d_j \beta_j$ , and that unfolding the recursive formula (15) produces the estimate of  $\|v_{sr+m}^{(r)}\|_{(1-\delta_r)R}$  as a sum of many terms every one of which contains as denominator a product of  $q$  divisors of the form  $d_{j_1} \beta_{j_1} \cdots d_{j_s} \beta_{j_s}$ , with some indexes  $j_1, \dots, j_s$  and some  $q$  to be found. This is what we call the *accumulation of small divisors*, and the problem is to identify the smallest one 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.

For given integers  $r < s$  let us consider the set

$$(16) \quad \mathcal{J}_{r,s} = \{I = \{j_1, \dots, j_{s-1}\} : j_m \in \{0, \dots, \min(r, s/2)\}, \mathbf{S}\}$$

where  $\mathbf{S}$  is the following

**Selection rule S:** every  $I \in \mathcal{J}_{r,s}$  satisfies  $I \triangleleft I_s^*$ , where

$$(17) \quad I_s^* = \left\{ \left\lfloor \frac{s}{s} \right\rfloor, \left\lfloor \frac{s}{s-1} \right\rfloor, \dots, \left\lfloor \frac{s}{2} \right\rfloor \right\} .$$

**Lemma 4:** For the set of indexes  $I_s^* = \{j_1, \dots, j_s\}$  the following statements hold true:

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

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

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

(ii) Let  $k$  be fixed and let  $m = \left\lfloor \frac{s}{k} \right\rfloor$ . Then we have  $s = mk + l$  for some  $0 \leq l < k$ . Therefore we also have

$$\left\lfloor \frac{mk+l}{m} \right\rfloor \geq k, \quad \left\lfloor \frac{mk+l}{m+1} \right\rfloor = \left\lfloor \frac{(m+1)k + (l-k)}{m+1} \right\rfloor < k,$$

namely we have  $\left\lfloor \frac{mk+l}{j} \right\rfloor \geq k$  if and only if  $j \leq \left\lfloor \frac{s}{k} \right\rfloor$ . On the other hand, by the same argument we have  $\left\lfloor \frac{mk+l}{j} \right\rfloor > k$  if and only if  $j \leq \left\lfloor \frac{s}{k+1} \right\rfloor$ . Thus we have  $\left\lfloor \frac{mk+l}{j} \right\rfloor = k$  if and only if

$$\left\lfloor \frac{s}{k+1} \right\rfloor < j \leq \left\lfloor \frac{s}{k} \right\rfloor .$$

There are exactly  $\left\lfloor \frac{s}{k} \right\rfloor - \left\lfloor \frac{s}{k+1} \right\rfloor$  such values of  $j$ , as claimed.

(iii) The case  $r = s$  is readily settled by reordering the indexes in increasing order as

$$\begin{aligned} \{r\} \cup I_r^* \cup I_r^* &= \left\{ \left\lfloor \frac{r}{r} \right\rfloor, \left\lfloor \frac{r}{r} \right\rfloor, \left\lfloor \frac{r}{r-1} \right\rfloor, \left\lfloor \frac{r}{r-1} \right\rfloor, \dots, \left\lfloor \frac{r}{2} \right\rfloor, \left\lfloor \frac{r}{2} \right\rfloor, \left\lfloor \frac{r}{1} \right\rfloor \right\} \\ I_{2r}^* &= \left\{ \left\lfloor \frac{2r}{2r} \right\rfloor, \left\lfloor \frac{2r}{2r-1} \right\rfloor, \left\lfloor \frac{2r}{2r-2} \right\rfloor, \left\lfloor \frac{2r}{2r-3} \right\rfloor, \dots, \left\lfloor \frac{2r}{4} \right\rfloor, \left\lfloor \frac{2r}{3} \right\rfloor, \left\lfloor \frac{2r}{2} \right\rfloor \right\} \end{aligned}$$

which makes the comparison straightforward. For  $r < s$  we need some further analysis. Let us define

$$M_k = \#\{j \in \{r\} \cup I_r^* \cup I_s^* : j \leq k\}, \quad N_k = \#\{j \in I_{r+s}^* : j \leq k\}, \quad 1 \leq k \leq \left\lfloor \frac{r+s}{2} \right\rfloor.$$

The upper limit for  $k$  is justified because no index exceeding  $\left\lfloor \frac{r+s}{2} \right\rfloor$  may appear in the sets above. Note also that in view of  $r < s$  we also have  $k < s$ . We claim that for  $1 \leq k \leq \left\lfloor \frac{r+s}{2} \right\rfloor$  we have

$$M_k = r + s - \left\lfloor \frac{r}{k+1} \right\rfloor - \left\lfloor \frac{s}{k+1} \right\rfloor, \quad N_k = r + s - \left\lfloor \frac{r+s}{k+1} \right\rfloor.$$

The second equality is a direct consequence of property (ii). For we easily calculate

$$N_k = \left( \left\lfloor \frac{r+s}{1} \right\rfloor - \left\lfloor \frac{r+s}{2} \right\rfloor \right) + \left( \left\lfloor \frac{r+s}{2} \right\rfloor - \left\lfloor \frac{r+s}{3} \right\rfloor \right) + \dots + \left( \left\lfloor \frac{r+s}{k} \right\rfloor - \left\lfloor \frac{r+s}{k+1} \right\rfloor \right)$$

which immediately gives the wanted equality. The first equality requires some attention, due to the extra index  $r$ . For  $k < r$  the index  $r$  is influential, and the formula is checked immediately by adding up the contributions of both  $I_r^*$  and  $I_s^*$ , which means that we just repeat the same calculation as for  $N_k$ . For  $k \geq r$  the formula is still correct because no index in  $\{r\} \cup I_r^*$  exceeds  $r$ , so that it contributes exactly  $r = r - \left\lfloor \frac{r}{k+1} \right\rfloor$  indexes, while  $s - \left\lfloor \frac{s}{k+1} \right\rfloor$  are contributed by  $I_s^*$ .

We claim now that for  $1 \leq r < s$  and  $1 \leq k \leq \left\lfloor \frac{r+s}{2} \right\rfloor$  we have  $N_k \leq M_k$ . This is a consequence of the inequality

$$\left\lfloor \frac{r}{k+1} \right\rfloor + \left\lfloor \frac{s}{k+1} \right\rfloor \leq \left\lfloor \frac{r+s}{k+1} \right\rfloor,$$

which is checked as follows. Let  $j = \left\lfloor \frac{r}{m} \right\rfloor$  and  $l = \left\lfloor \frac{s}{m} \right\rfloor$ , so that we can write  $r = jm + p$  and  $s = lm + q$  with  $0 \leq p < m$  e  $0 \leq q < m$ . Then we calculate

$$\left\lfloor \frac{r+s}{m} \right\rfloor = \left\lfloor \frac{(j+l)m + p + q}{m} \right\rfloor = j + l + \left\lfloor \frac{p+q}{m} \right\rfloor \geq j + l,$$

which is the wanted inequality.

Using  $N_k \leq M_k$  it is now an easy matter to complete the proof, by just reordering both the set of indexes  $r \cup I_r^* \cup I_s^*$  and  $I_{r+s}^*$  in increasing order. For, in view of  $M_1 \geq N_1$  every index 1 in  $r \cup I_r^* \cup I_s^*$  has a corresponding index in  $I_{r+s}^*$  which is at least 1; in view of  $M_2 \geq N_2$  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  $k = \left\lfloor \frac{r+s}{2} \right\rfloor$ . We conclude that  $r \cup I_r^* \cup I_s^* \prec I_{r+s}^*$ , as claimed. *Q.E.D.*

**Lemma 5:** 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. This is seen as follows. Looking at the recursive estimate (15) one sees that a generic set  $\tilde{I}$  of indexes for the divisors  $d_j \beta_j$  in  $\|v_s^{(r)}\|_{(1-\delta_r)R}$  is constructed as

$$\tilde{I} = \underbrace{\{r\} \cup \dots \cup \{r\}}_{p \text{ times}} \cup I_1 \cup \dots \cup I_p \cup I' ,$$

$$1 \leq p < s, \quad I_1 \in \mathcal{J}_{r-1,r}, \dots, I_p \in \mathcal{J}_{r-1,r}, \quad I' \in \mathcal{J}_{r-1,(s-p)r+m} ;$$

For, the index  $r$  appears explicitly in the divisor  $d_r \beta_r$ , the sets of indexes  $I_1, \dots, I_p$  are contributed by  $\|v_r^{(r-1)}\|_{(1-\delta_{r-1})R}$  raised to the power  $p$ , and the set of indexes  $I'$  is contributed by  $\|v_{(s-p)r+m}^{(r-1)}\|_{(1-\delta_{r-1})R}$ . The definition above of  $\tilde{I}$  can be replaced with the simpler one

$$\tilde{I} = \{r\} \cup I \cup I' , \quad I \in \mathcal{J}_{r-1,r}, \quad I' \in \mathcal{J}_{r,s} .$$

**Proof of lemma 5.** (i) is immediately checked in view of the definition of  $\mathcal{J}_{r,s}$ .

(ii) We have  $\#(\tilde{I}) < r+s$  in view of  $\#(\{r\} \cup I \cup I') = 1 + \#(I) + \#(I') = 1 + r - 1 + s - 1 = r + s - 1$ . Let now  $j \in \tilde{I}$ . Then 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 4. Q.E.D.

We come now to consider the accumulation of small divisors. Recall the definition (3) of the sequence  $\beta_r$ , and that we have introduced a sequence  $d_1, d_2, \dots$  of positive numbers satisfying  $\sum_{j \geq 1} d_j = d < \infty$ . We associate to the sets of indexes  $\mathcal{J}_{r,s}$  the sequence of positive real numbers  $T_{r,s}$  defined as

$$(18) \quad T_{0,s} = 1, \quad T_{r,s} = \max_{I \in \mathcal{J}_{r,s}} \prod_{j \in I} \frac{1}{d_j \beta_j}, \quad 0 < r < s .$$

**Lemma 6:** *The sequence  $T_{r,s}$  satisfies the following properties for  $1 < r < s$ :*

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

**Proof.** (i) From property (i) of lemma 5 we readily get

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

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

(ii) Compute

$$\begin{aligned}
\frac{1}{d_r \beta_r} T_{r-1,r} T_{r,s} &= \frac{1}{d_r \beta_r} \max_{I \in \mathcal{J}_{r-1,r}} \prod_{j \in I} \frac{1}{d_j \beta_j} \max_{I' \in \mathcal{J}_{r,s}} \prod_{j' \in I'} \frac{1}{d_{j'} \beta_{j'}} \\
&= \max_{I \in \mathcal{J}_{r-1,r}} \max_{I' \in \mathcal{J}_{r,s}} \prod_{j \in \{r\} \cup I \cup I'} \frac{1}{d_j \beta_j} \\
&\leq \max_{I \in \mathcal{J}_{r,r+s}} \prod_{j \in I} \frac{1}{d_j \beta_j} = T_{r,r+s} ,
\end{aligned}$$

where in the inequality of the last line property (ii) of lemma 5 has been used. *Q.E.D.*

The final estimate makes use of the definition (4) of the sequence  $\alpha_r$  and of condition  $\tau$ . We also should make a choice for the sequence  $d_r$  that enters the definition of the sequence  $T_{r,s}$ . Recalling that the condition  $\sum_{j \geq 1} d_j = d < 1$  must be satisfied we make the simple choice

$$(19) \quad d_r = \frac{6d}{\pi^2} \cdot \frac{1}{r^2} .$$

leaving  $d$  still arbitrary.

**Lemma 7:** *Let  $\lambda$  satisfy condition  $\tau$  and the sequence  $d_r$  be defined as in (19). Then the sequence  $T_{r,s}$  is bounded by*

$$T_{r,s} \leq A^s e^{s\Gamma} , \quad \frac{1}{d_s \beta_s} T_{r,s} \leq A^s e^{s\Gamma}$$

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

**Proof.** In view of  $d_s \beta_s \leq 1$  it is enough to prove the second inequality. We use the definition (4) of the sequence  $\alpha_r$ , the property (ii) of lemma 5 and the selection rule S. We readily get

$$T_{r,s} = \frac{1}{d_s \beta_s} \max_{I \in \mathcal{J}_{r,s}} \prod_{j \in I} \frac{1}{d_j \beta_j} \leq \frac{1}{d_s \beta_s} \max_{I \in \mathcal{J}_{r,s}} \prod_{j \in I} \frac{1}{d_j \alpha_j} \leq \prod_{j \in \{s\} \cup I_s^*} \frac{1}{d_j \alpha_j} .$$

By property (ii) of lemma 4 the latter product is evaluated as

$$\prod_{j \in \{s\} \cup I_s^*} \frac{1}{d_j \alpha_j} = [(d_1 \alpha_1)^{q_1} \cdot \dots \cdot (d_{\lfloor s/2 \rfloor} \alpha_{\lfloor s/2 \rfloor})^{q_{\lfloor s/2 \rfloor}} (d_s \alpha_s)^{q_s}]^{-1} ,$$

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  $k$ . Thus we have

$$\begin{aligned}
\ln T_{r,s} &\leq - \sum_{k=1}^s \left( \left\lfloor \frac{s}{k} \right\rfloor - \left\lfloor \frac{s}{k+1} \right\rfloor \right) (\ln \alpha_k + \ln d_k) \\
&\leq -s \sum_{k \geq 1} \frac{\ln \alpha_k + \ln d_k}{k(k+1)} = s \left( \Gamma - \sum_{k \geq 1} \frac{\ln d_k}{k(k+1)} \right) .
\end{aligned}$$

Replacing the value of  $d_k$  as given by (19) we have

$$-\sum_{k \geq 1} \frac{\ln d_k}{k(k+1)} = \sum_{k \geq 1} \frac{1}{k(k+1)} \left( \ln \frac{\pi^2}{6d} + 2 \ln k \right) = a ,$$

where  $a$  denotes the sum of the series, which clearly converges. Remark that the latter quantity depends only on the choice of the sequence  $d_k$ , and not on  $\lambda$ ; this is part of the statement. We conclude

$$\ln T_{r,s} < s(\Gamma + a) ,$$

and the claim follows by just setting  $A = e^a$ .

*Q.E.D.*

## 4. Proof of the main theorem

We come finally to completing the proof of the theorem. This is now matter of finding bounds for the sequence  $X_r$  of generating vector fields that give the system the normal form and then proving the convergence of the sequence of transformations.

### 4.1 Estimates for the generating vector fields

In view of the analyticity of the vector field  $v$  in the r.h.s. of eq. (1) the inequality

$$(20) \quad \|v_s^{(0)}\|_R \leq b^{s-1} E , \quad b \geq 0 , \quad E > 0 .$$

holds true for some constants  $E$  and  $b$ . Going back to the recursive formula of lemma 3 we prove the following

**Lemma 8:** *With the hypothesis (20) we have*

$$(21) \quad \|v_{sr+m}^{(r)}\|_{(1-\delta_r)R} \leq \nu_{r,sr+m} T_{r,sr+m} C^{sr+m-1} E$$

where

$$(22) \quad C = \frac{2eE}{Q} + b .$$

is a constant not depending on  $\lambda$ ,  $\nu_{r,sr+m}$  is a double indexed sequence defined as

$$(23) \quad \nu_{0,s} = 1 , \quad \nu_{r,sr+m} = \sum_{p=0}^{s-1} \nu_{r-1,r}^p \nu_{r-1,(s-p)r+m}$$

for  $r > 0$  ,  $s \geq 1$  ,  $0 \leq m < r$   
 excluding  $(s = 1 , m = 0)$  ,

and  $T_{r,s}$  is the sequence defined by (18).

**Proof.** For  $r = 0$  the claim is trivially true for in view of  $\nu_{0,s} = 1$  and  $T_{0,s} = 0$  for  $s \geq 1$ . By induction, let  $r > 0$  and replace (21) in (15), thus getting

$$(24) \quad \begin{aligned} & \|v_{sr+m}^{(r)}\|_{(1-\delta_r)R} \\ & \leq \sum_{p=0}^{s-1} \nu_{r-1,r}^p \nu_{r-1,(s-p)r+m} \left( \frac{T_{r-1,r}}{d_r \beta_r} \right)^p T_{r-1,(s-p)r+m} \left( \frac{2eEC^{r-1}}{Q} \right)^p C^{(s-p)r+m-1} E \\ & \leq \sum_{p=0}^{s-1} \nu_{r-1,r}^p \nu_{r-1,(s-p)r+m} T_{r,sr+m} C^{sr+m-1} E . \end{aligned}$$

Let us check the latter inequality. For  $r = 1$  it is true in view of (18), since  $T_{0,s} = 1$  e  $T_{1,s} = \max\left(\frac{1}{d_1 \beta_1}, \dots, \frac{1}{(d_1 \beta_1)^{s-1}}\right)$ . For  $r > 1$  we repeatedly apply lemma 6, and calculate

$$\begin{aligned} \left( \frac{T_{r-1,r}}{d_r \beta_r} \right)^p T_{r-1,(s-p)r+m} &= \left( \frac{T_{r-1,r}}{d_r \beta_r} \right)^{p-1} \frac{T_{r-1,r}}{d_r \beta_r} T_{r,(s-p)r+m} \\ &\leq \left( \frac{T_{r-1,r}}{d_r \beta_r} \right)^{p-2} \frac{T_{r-1,r}}{d_r \beta_r} T_{r,(s-p+1)r+m} \leq \dots \leq \frac{T_{r-1,r}}{d_r \beta_r} T_{r,(s-1)r+m} \leq T_{r,sr+m} . \end{aligned}$$

Thus we see that (24) holds true provided the sequence  $\nu_{r,s}$  is defined as in (23). *Q.E.D.*

The last lemma gives a bound on the sequence of vector fields  $X_r$ .

**Lemma 9:** *If condition  $\tau$  is fulfilled then the sequence  $X_r$  of generating vector fields is bounded by*

$$(25) \quad \|X_r\|_{(1-d)R} \leq \frac{1}{r^2} D^{r-1} e^{r\Gamma} F$$

where  $D$  and  $F$  are positive constants not depending on  $\lambda$ .

**Proof.** In view of (14) and (21) we have

$$\|X_r\|_{(1-d)R} \leq \frac{1}{\beta_r} \|v_r^{(r-1)}\|_{(1-d)R} \leq \frac{1}{\beta_r} \nu_{r-1,r} T_{r-1,r} C^{r-1} E$$

where  $C$  is the constant given by (22). According to lemma 7 we also have

$$\frac{1}{\beta_r} T_{r-1,r} \leq d_r A^r e^{r\Gamma} = \frac{6d}{\pi^2 r^2} A^r e^{r\Gamma}$$

where the constant  $A$  is evaluated in the proof of the lemma and is independent of  $\lambda$ . It remains to evaluate the constants  $\nu_{r-1,r}$  defined by the sequence (23). This requires some boring calculation that is deferred to appendix A.3, where we prove that  $\nu_{r-1,r} \leq 4^{r-1}$ . Collecting all these informations the claim follows by just setting

$$D = 4AC , \quad F = \frac{6d}{\pi^2} AE .$$

*Q.E.D.*

## 4.2 Completion of the proof

Having established the estimate of lemma 9 on the sequence of generating function 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.

Stating the problem in general terms the situation to be dealt with is the following. An infinite sequence  $\{X_r\}_{r \geq 1}$  of generating vector fields is given, with  $X_r$  homogeneous polynomial of degree  $r+1$  satisfying  $\|X_r\|_R \leq \beta^{r-1}G$  for some real vector  $R$  with positive components and some positive  $\beta$  and  $G$ . Define a corresponding sequence of near the identity transformations  $x^{(r-1)} = \exp(L_{X_r})x^{(r)}$ . By composition one also constructs a sequence  $\{\mathcal{C}^{(r)}\}_{r \geq 0}$  of transformations  $x^{(0)} = \mathcal{C}^{(r)}x^{(r)}$  recursively defined as

$$\mathcal{C}^{(0)} = \text{I} , \quad \mathcal{C}^{(r)} = \exp(L_{X_r}) \circ \mathcal{C}^{(r-1)} ,$$

$\text{I}$  being the identity operator. The problem is to prove the following statements.

- (i) Every near the identity canonical transformation defined via the exponential operator  $\exp(L_{X_r})$  is expressed as a power series which is convergent in a polydisk  $\Delta_{\varrho R}$  for some positive  $\varrho$ .
- (ii) For any function  $f(x^{(r-1)})$  analytic in  $\Delta_{\varrho R}$  the transformed function is analytic in the same polydisk, and moreover

$$f(x^{(r-1)}) \Big|_{x^{(r-1)} = \exp(L_{X_r})x^{(r)}} = [\exp(L_{X_r})f](x^{(r)}) .$$

- (iii) The sequence  $\{\mathcal{C}^{(r)}\}_{r \geq 0}$  of transformations converges for  $r \rightarrow \infty$  to a transformation  $\mathcal{C}^{(\infty)}$  which is analytic in a polydisk  $\Delta_{(1-d)\varrho R}$  for some positive  $d < 1/2$ .
- (iv) For any function  $f$  analytic in  $\Delta_{\varrho R}$  the sequence recursively defined as  $f^{(0)} = f$ ,  $f^{(r)} = \exp(L_{X_r})f^{(r-1)}$  converges for  $r \rightarrow \infty$  to a function  $f^{(\infty)}$  that is analytic in  $\Delta_{(1-d)\varrho R}$ , and moreover one has

$$f^{(\infty)} = f \circ \mathcal{C}^{(\infty)} .$$

The statement (i) actually reduces to Cauchy's proof of the existence and uniqueness of the local solution of an analytic system of differential equations. The statement (ii) actually claims that the substitution of variables in a function  $f$  may be effectively replaced by the application of the exponential operator to  $f$ ; this is indeed the basis of the algorithm for constructing the normal form used in sect. 2. A detailed proof of both these statements may be found, e.g., in [12]; however, the reader may be able to reconstruct the proof by following the hints in [9].

The proof of (iii) rests on the following remarks. In the polydisk  $\Delta_{\varrho R}$  one has  $|X_r(x)| \leq \varrho^{r+1} \|X_r\|_{\varrho R}$ ; this, in turn, implies that  $|x^{(r)} - x^{(r-1)}| \sim \beta^{r-1} \varrho^{r+1}$ . The geometric bound on the latter quantities implies that  $\sum_{r \geq 1} |x^{(r)} - x^{(r-1)}|$  is bounded by a geometric series, i.e., converges for  $\varrho$  small enough. Thus, the claim follows from Weierstrass theorem. Finally, the statement (iv) follows from (ii) being true for all  $r > 0$ , which implies that both sequences  $f^{(r)} = \mathcal{C}^{(r)}f$  and  $f \circ \mathcal{C}^{(r)}$  converge to the same limit. This concludes the proof of theorem 1.

## 5. Relations between condition $\tau$ and Bruno's function

In this section we show the equivalence between Bruno's condition and condition  $\tau$ . Furthermore we discuss the relations between our condition  $\tau$  and the Bruno function.

### 5.1 Comparison between condition $\tau$ and the Bruno condition

In [5] the convergence of the normalizing transformation  $\psi$  is proved under the condition

$$(26) \quad \mathbb{B} = - \sum_{r \geq 1} \frac{1}{2^r} \log \alpha_{2^r-1} < +\infty .$$

Moreover  $\psi$  is shown to be analytic in a polydisk  $|x_i| < A^{-1}b^{-1}e^{-2\mathbb{B}}$ ,  $i = 1, \dots, n$ . In this section we will prove that condition  $\tau$  is equivalent to (26) and that for all nonresonant  $\Lambda$  one has

$$(27) \quad \Gamma < \mathbb{B} < 2 \left( \Gamma - \frac{\log \alpha_1}{2} \right) .$$

From the estimate above and Theorem 1 it follows that  $\psi$  is analytic in a polydisk of radius at least equal to  $A^{-1}b^{-1}e^{-\mathbb{B}}$ , as we claimed in the Introduction.

Let us check (27). From the definition one has

$$\alpha_{2^r-1} \geq \alpha_{2^{r-1}+1} \geq \dots \geq \alpha_{2^r-1}$$

thus

$$- \sum_{k=2^{r-1}}^{2^r-1} \frac{\log \alpha_k}{k(k+1)} \leq - \log \alpha_{2^r-1} \sum_{k=2^{r-1}}^{2^r-1} \frac{1}{k(k+1)} = - \frac{\log \alpha_{2^r-1}}{2^r} .$$

This obviously implies the lower bound in (27) since

$$\Gamma = - \sum_{k \geq 1} \frac{\log \alpha_k}{k(k+1)} = - \sum_{r \geq 1} \sum_{k=2^{r-1}}^{2^r-1} \frac{\log \alpha_k}{k(k+1)} < - \sum_{r \geq 1} \frac{\log \alpha_{2^r-1}}{2^r} = \mathbb{B} .$$

For the upper bound one simply observes that

$$\alpha_{2^r-1} \geq \dots \geq \alpha_{2^{r+1}-1}$$

thus

$$- \frac{\log \alpha_{2^r-1}}{2^r} = -2 \log \alpha_{2^r-1} \sum_{k=2^r}^{2^{r+1}-1} \frac{1}{k(k+1)} \leq -2 \sum_{k=2^r}^{2^{r+1}-1} \frac{\log \alpha_k}{k(k+1)}$$

which leads to

$$- \sum_{r \geq 1} \frac{1}{2^r} \log \alpha_{2^r-1} \leq -2 \sum_{r \geq 1} \sum_{k=2^r}^{2^{r+1}-1} \frac{\log \alpha_k}{k(k+1)} = -2 \left( \Gamma - \frac{\log \alpha_1}{2} \right) .$$

## 5.2 Comparison between condition $\tau$ and the Bruno function

Let  $n = 2$ ,  $\lambda = (1, x)$  where  $x$  is an irrational number. By means of the best approximation property of continued fractions the Bruno condition (26) can be equivalently recasted as the requirement of  $x$  being a *Bruno number*, namely such that the infinite sum

$$(28) \quad \sum_{n=0}^{\infty} q_n^{-1} \log q_{n+1} < +\infty,$$

where  $(p_n/q_n)_{n \geq 0}$  is the sequence of the convergents of its continued fraction expansion (we refer to the appendix A.4 for a short introduction to continued fractions and their relation with diophantine approximation problems). By introducing (as in [29], [19]) the *Bruno function*  $B : \mathbb{R} \setminus \mathbb{Q} \rightarrow (0, +\infty]$ ,

$$(29) \quad B(x) := \sum_{n=0}^{\infty} b_{n-1}(x) \log x_n^{-1}$$

the set of Bruno numbers is characterized as the subset of  $\mathbb{R} \setminus \mathbb{Q}$  on which  $B$  is finite, since it is easy to show that there exists a universal constant  $C > 0$  such that for all  $x \in \mathbb{R} \setminus \mathbb{Q}$  one has

$$\left| B(x) - \sum_{n=0}^{\infty} \frac{\log q_{n+1}}{q_n} \right| \leq C.$$

All diophantine numbers are Bruno numbers but also “many” Liouville numbers are Bruno numbers: for example  $\sum_{n \geq 1} 10^{-n!}$  is a Bruno number.

The main advantage of the Bruno function over the series (28) is that it satisfies a remarkable functional equation under the action of the generators of the modular group:

$$(30) \quad \begin{aligned} B(x) &= B(x+1), \quad \forall x \in \mathbb{R} \setminus \mathbb{Q} \\ B(x) &= -\log x + xB\left(\frac{1}{x}\right), \quad x \in \mathbb{R} \setminus \mathbb{Q} \cap (0, 1) \end{aligned}$$

This makes clear that the set of Bruno numbers is  $\mathrm{SL}(2, \mathbb{Z})$ -invariant. Moreover, since quadratic irrationals have an eventually periodic continued fraction expansion, for each of them one can compute the Bruno function *exactly* with finitely many iterations of (30). Thus  $B$  is known *exactly* on a countable but dense set of irrationals.

In order to compare the Bruno function with condition  $\tau$  it is convenient to introduce the function

$$\omega_r(x) = \min_{1 \leq q \leq r, p \in \mathbb{Z}} |p - qx|$$

so that by the best approximation property of continued fractions one has  $\omega_r(x) = \omega_{q_k}(x)$  whenever  $q_k \leq r < q_{k+1}$ . From the definitions it is immediate to check that

$$\beta_r \geq \omega_r$$

for all  $r \geq 0$  thus also  $\alpha_k \geq \omega_k$  for all  $k \geq 1$  and

$$(31) \quad \Gamma \leq \tilde{\Gamma} = - \sum_{k \geq 1} \frac{\omega_k}{k(k+1)}.$$

On the other hand one can rewrite  $\tilde{\Gamma}$  as follows

$$\tilde{\Gamma} = - \sum_{k \geq 1} \sum_{j=q_k}^{q_{k+1}-1} \left( \frac{1}{j} - \frac{1}{j+1} \right) \log |q_k x - p_k| = - \sum_{k \geq 1} \left( \frac{1}{q_k} - \frac{1}{q_{k+1}} \right) \log |q_k x - p_k|.$$

Rearranging terms in the sum one gets

$$\tilde{\Gamma} = - \sum_{k \geq 1} \frac{1}{q_k} \log \left| \frac{q_{k-1} x - p_{k-1}}{q_k x - p_k} \right|,$$

which differs from the Bruno function only in the use of  $\frac{1}{q_k}$  instead of  $b_{k-1}$ : recall however that for all  $k \geq 1$  one has  $\frac{1}{2} < b_{k-1} q_k < 1$

## A. Technical proofs

We collect in this appendix a number of technical details that are needed in order to complete the proofs. We also include a short discussion about continued fractions and Bruno series.

### A.1 Lie series algorithm for the normal form

In this section we give a full justification of the algorithm for giving the system (1) a normal form using Lie series. The problem is to find explicit expressions for the transformed vector field  $\exp(L_{X_r})(\Lambda x + v_r^{(r-1)} + \dots)$  by suitably reordering all terms. The linearity of the operator  $\exp(L_{X_r})$  will be used.

Consider first  $\Lambda x + v_r^{(r-1)}$  and calculate

$$\begin{aligned} \exp(L_{X_r})(\Lambda x + v_r^{(r-1)}) &= \Lambda x + L_{X_r}(\Lambda x) + \sum_{s \geq 2} \frac{1}{s!} L_{X_r}^s(\Lambda x) \\ &\quad + v_r^{(r-1)} + \sum_{s \geq 1} \frac{1}{s!} L_{X_r}^s v_r^{(r-1)}. \end{aligned}$$

Collecting terms of polynomial degree  $r$  we get  $L_{X_r}(\Lambda x) + v_r^{(r-1)}$ . Since we want to remove such terms from the normal form we immediately get for  $X_r$  the equation  $L_{X_r}(\Lambda x) + v_r^{(r-1)} = 0$ . This is nothing but the homological equation (8). For, we have  $L_{X_r}(\Lambda x) = -L_{\Lambda x} X_r$ , and we just denote  $D_\lambda = L_{\Lambda x}$ . Then we rearrange all terms of degree higher than  $r$  by calculating

$$\begin{aligned} &\sum_{s \geq 2} \frac{1}{s!} L_{X_r}^s(\Lambda x) + \sum_{s \geq 1} \frac{1}{s!} L_{X_r}^s v_r^{(r-1)} \\ &= \sum_{s \geq 2} \frac{1}{(s-1)!} L_{X_r}^{s-1} \left[ \frac{1}{s} \left( L_{X_r}(\Lambda x) + v_r^{(r-1)} \right) + \frac{s-1}{s} v_r^{(r-1)} \right] \\ &= \sum_{s \geq 2} \frac{s-1}{s!} L_{X_r}^{s-1} v_r^{(r-1)}, \end{aligned}$$

where the homological equation has been used in order to remove the first term in square brackets in the second line. In the latter sum the term  $L_{X_r}^{s-1} v_r^{(r-1)}$  is a homogeneous polynomial of degree  $sr + 1$ , and we include it in the expression for  $v_{sr}^{(r)}$  in the second of (9).

The rest of the expansion comes from the transformation of  $v_s^{(r-1)}$  with  $s > r$ . It is convenient to introduce two separate indexes by setting  $s = lr + m$  with  $l \geq 1$  e  $0 \leq m < r$  the case  $l = 1, m = 0$  must be removed because  $v_r^{(r)}$  never appears. Thus we get

$$\exp(L_{X_r}) v_{lr+m}^{(r-1)} = \sum_{p \geq 0} \frac{1}{p!} L_{X_r}^p v_{lr+m}^{(r-1)}$$

where  $L_{X_r}^p v_{lr+m}^{(r-1)}$  is a homogeneous polynomial of degree  $(p+l)r + m$ . By collecting all homogeneous terms with  $m = 0, l \geq 2$  and  $p+l = s \geq 2$  we get  $\sum_{p=0}^{s-2} \frac{1}{p!} L_{X_r}^p v_{(s-p)r}^{(r-1)}$ , that we add to  $v_{sr}^{(r)}$  in the second of (9). Similarly, collecting the homogeneous terms with  $0 < m < r, l \geq 1$  and  $p+l = s \geq 1$  we get  $\sum_{p=0}^{s-1} \frac{1}{p!} L_{X_r}^p v_{(s-p)r+m}^{(r-1)}$ , that we add to  $v_{sr+m}^{(r)}$  in the first of (9). The latter case occurs only for  $r > 1$ . This exhausts all of the contributions to the transformed field. Thus the formal algorithm is justified.

## A.2 Generalized Cauchy estimates

We give here the proof of lemma 1. Write the components of the vector field as

$$X_j = \sum_k c_{j,k} x^k, \quad v_j = \sum_{k'} v_{j,k'} x^{k'}.$$

Then calculate the components of the Lie derivative as

$$(L_X v)_j = \sum_{l=1}^n \sum_{k,k'} \left( c_{l,k} x^k k'_l v_{j,k} \frac{x^{k'}}{x_l} - v_{l,k'} x^{k'} k_l c_{j,k} \frac{x^k}{x_l} \right).$$

Calculating the norm of the latter expression we get

$$\begin{aligned} \|(L_X v)_j\|_{(1-d)R} &= \sum_{l=1}^n \sum_k |c_{l,k}| (1-d)^{|k|} R^k \sum_{k'} k'_l |v_{j,k}| (1-d)^{|k'|-1} \frac{R^{k'}}{R_l} \\ &\quad + \sum_{k'} |v_{l,k'}| (1-d)^{|k'|} R^{k'} \sum_k k_l |c_{j,k}| (1-d)^{|k|-1} \frac{R^k}{R_l}. \end{aligned}$$

Considering only the first part we have

$$\begin{aligned} \sum_{l=1}^n \sum_k |c_{l,k}| (1-d)^{|k|} R^k \sum_{k'} k'_l |v_{j,k}| (1-d)^{|k'|-1} \frac{R^{k'}}{R_l} \\ \leq \frac{1}{Q} \sum_{l=1}^n \|X\|_{(1-d)R} \sum_{k'} k'_l |v_{j,k}| (1-d)^{|k'|-1} R^{k'}, \end{aligned}$$

where the definition of the norm  $\|X\|_{(1-d)R}$  has been used. In the latter expression only  $k'_l$  depends on the index  $l$ , and we are allowed to substitute  $\sum_{l=1}^n k'_l = |k'|$ . Moreover we use the inequality<sup>†</sup>

$$|k'|(1-d)^{|k'|-1} = |k'|((1-d') - (d-d'))^{|k'|-1} < \frac{(1-d')^{|k'|}}{d-d'} ,$$

so that the expression above is bounded by

$$\frac{1}{(d-d')Q} \|X\|_R \|v_j\|_{(1-d')R} .$$

A similar calculation leads to

$$\sum_{k'} |v_{l,k'}|(1-d)^{|k'|} R^{k'} \sum_k k_l |c_{j,k}| (1-d)^{|k|-1} \frac{R^k}{R_l} \leq \frac{1}{dQ} \|v\|_{(1-d')R} \|X_j\|_R ,$$

and collecting all the estimates and using the definition of the norm of a vector field we get

$$(32) \quad \|L_X v\|_{(1-d)R} \leq \frac{2}{(d-d')Q} \|X\|_R \|v\|_{(1-d')R} .$$

This coincides with (12) for  $s = 1$ .

Let now  $s > 1$  and set  $\delta = (d-d')/s$ . By applying (32)  $s$  times we calculate

$$\begin{aligned} \|L_X^2 v\|_{(1-d'-2\delta)R} &\leq \frac{2}{\delta Q} \|X\|_R \|L_X v\|_{(1-d'-\delta)R} \leq \left(\frac{2}{\delta Q}\right)^2 \|X\|_R^2 \|v\|_{(1-d')R} , \\ \|L_X^3 v\|_{(1-d'-3\delta)R} &\leq \frac{2}{\delta Q} \|X\|_R \|L_X^2 v\|_{(1-d'-2\delta)R} \leq \left(\frac{2}{\delta Q}\right)^3 \|X\|_R^3 \|v\|_{(1-d')R} , \\ &\vdots \\ \|L_X^s v\|_{(1-d'-s\delta)R} &\leq \left(\frac{2}{\delta Q}\right)^s \|X\|_R^s \|v\|_{(1-d')R} . \end{aligned}$$

Now we replace  $\delta = (d-d')/s$  in the latter line and use the inequality  $s^s \leq s!e^{s-1}$ , thus getting (12). This concludes the proof of lemma 1

### A.3 Estimate of the sequence (23)

We first introduce a more convenient sequence  $\eta_{r,s}$  defined as

$$(33) \quad \begin{aligned} \eta_{0,s} &= 1 , \\ \eta_{r,s} &= \sum_{0 < p_1 + 2p_2 + \dots + rp_r < s} \eta_{r-1,r}^{p_r} \eta_{r-2,r-1}^{p_{r-1}} \cdots \eta_{0,1}^{p_1} \eta_{0,s-rp_r-\dots-2p_2-p_1} , \quad s > r . \end{aligned}$$

---

<sup>†</sup> Use the general inequality  $m(a-x)^{m-1} < \frac{a^m}{x}$  for  $0 \leq x \leq a$  and  $m \geq 1$ . This is easily checked by computing the maximum of the function  $x(a-x)^{m-1}$  over the interval  $0 \leq x \leq a$ .



For, if  $m \geq r - m$  the sum can be truncated because only  $q_1, \dots, q_{r-m}$  can be non null in view of  $q_1 + 2q_2 + \dots + mq_m < r - m$ ; if instead  $m < r - m$  then we add some extra positive terms, so that the inequality holds true. Thus we have

$$\eta_{r-1,r} \leq \sum_{m=1}^{r-1} \eta_{m-1,m} \eta_{r-m-1,r-m} ,$$

so that  $\eta_{r-1,r} \leq \mu_{r-1}$  in view of (34).

The final step is to estimate the sequence (34). This is the well known Catalan's sequence, which is explicitly determined as

$$(35) \quad \mu_r = \frac{2^{r-1}(2r-3)!!}{r!} \leq 4^{r-1} ,$$

where the notation  $(2n+1)!! = 1 \cdot 3 \cdot \dots \cdot (2n+1)$  for the so called semifactorial has been used.

#### A.4 Arithmetical tools and Bruno series

We recall here some elementary facts about classical continued fractions. We refer the reader to the classical books [13] and [15] for more details. Continued fractions arise naturally constructing the symbolic dynamics of the Gauss map (as well as for the linear flow on the two-dimensional torus or the geodesic flow on the modular surface). Here we will consider the iteration of the Gauss map  $\mathcal{G} : (0, 1) \mapsto [0, 1]$  defined by  $\mathcal{G}(x) = \{x^{-1}\} = x^{-1} - [x^{-1}]$  where  $[x]$  and  $\{x\}$  respectively denote the integer and the fractional part of  $x$ . This map is piecewise analytic with inverse branches  $T_n(x) = \frac{1}{n+x}$ ,  $T_n = \mathcal{G}^{-1}$  on the interval  $(\frac{1}{n+1}, \frac{1}{n})$ . Given  $x \in \mathbb{R} \setminus \mathbb{Q}$  we set  $x_0 = x - [x]$ ,  $a_0 = [x]$ , then one obviously has  $x = a_0 + x_0$ . We now define inductively for all  $n \geq 0$   $x_{n+1} = \mathcal{G}(x_n)$ ,  $a_{n+1} = [x_n^{-1}] \geq 1$ , thus  $x_n = T_{a_{n+1}}(x_{n+1})$ . Therefore we have

$$x = a_0 + T_{a_1}(x_1) = \dots = a_0 + T_{a_1} \circ \dots \circ T_{a_n}(x_n) = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \dots + \frac{1}{a_n + x_n}}} .$$

We will use the short notation  $x = [a_0, a_1, \dots, a_n, \dots]$  for the infinite fraction. The  $n$ th-convergent is then the rational number corresponding to the finite fraction  $\frac{p_n}{q_n} = [a_0, a_1, \dots, a_n]$ .

The numerators  $p_n$  and denominators  $q_n$  are recursively determined for all  $n \geq 0$  by

$$p_n = a_n p_{n-1} + p_{n-2} , q_n = a_n q_{n-1} + q_{n-2} ,$$

with the initial conditions  $p_{-1} = q_{-2} = 1$  ,  $p_{-2} = q_{-1} = 0$ . Note that  $q_n p_{n-1} - p_n q_{n-1} = (-1)^n$  and  $q_{n+1} > q_n > 0$  for all  $n \geq 0$ . The sequence of the numerators  $p_n$  has the same constant sign of  $x$ .

For all  $n \geq 0$  one also has

$$x = \frac{p_n + p_{n-1} x_n}{q_n + q_{n-1} x_n} , \quad x_n = -\frac{q_n x - p_n}{q_{n-1} x - p_{n-1}} ,$$

thus for all  $k \geq 0$  and for all  $x \in \mathbb{R} \setminus \mathbb{Q}$  one has  $\frac{p_{2k}}{q_{2k}} < x < \frac{p_{2k+1}}{q_{2k+1}}$ .

Let

$$b_n(x) = \prod_{i=0}^n x_i = (-1)^n (q_n x - p_n) \quad \text{for } n \geq 0, \quad \text{and } b_{-1}(x) = 1$$

so that

$$x_n = b_n b_{n-1}^{-1}, \quad \text{and } b_{n-2} = a_n b_{n-1} + b_n,$$

and let

$$G = \frac{\sqrt{5} + 1}{2}, \quad g = G^{-1} = \frac{\sqrt{5} - 1}{2}.$$

Using all the formulas above it is easy to show that for all  $x \in \mathbb{R} \setminus \mathbb{Q}$  and for all  $n \geq 1$  one has

- (i)  $|q_n x - p_n| = \frac{1}{q_{n+1} + q_n x_{n+1}}$ , so that  $\frac{1}{2} < b_n q_{n+1} < 1$ ;
- (ii)  $b_n \leq g^n$  and  $q_n \geq \frac{1}{2} G^{n-1}$ .

Note that from (ii) it follows that the two series  $\sum_{k=0}^{\infty} \frac{\log q_k}{q_k}$  and  $\sum_{k=0}^{\infty} \frac{1}{q_k}$  are always convergent and their sum is bounded by a universal constant.

The intimate connection between the modular group and the Gauss map appears also through the fact that two points  $x, y \in \mathbb{R} \setminus \mathbb{Q}$  have the same  $\text{SL}(2, \mathbb{Z})$ -orbit if and only if  $x = [a_0, a_1, \dots, a_m, c_0, c_1, \dots]$  and  $y = [b_0, b_1, \dots, b_n, c_0, c_1, \dots]$ .

The importance of continued fractions in the theory of diophantine approximation comes largely from the best approximation theorem ([13], respectively Theorems 182, p. 151 and 184, p. 153): for all irrational numbers  $x$  let  $p_n/q_n$  denote its  $n$ -th convergent. If  $0 < q < q_{n+1}$  then  $|qx - p| \geq |q_n x - p_n|$  for all  $p \in \mathbb{Z}$  and equality can occur only if  $q = q_n, p = p_n$ . Moreover if  $p/q$  is a rational number such that  $\left| x - \frac{p}{q} \right| < \frac{1}{2q^2}$  then  $\frac{p}{q}$  is necessarily a convergent of  $x$ .

Let  $\gamma > 0$  and  $\mu \geq 0$  be two real numbers: an irrational number  $x \in \mathbb{R} \setminus \mathbb{Q}$  is diophantine of exponent  $\mu$  and constant  $\gamma$  if and only if for all  $p, q \in \mathbb{Z}, q > 0$ , one has  $\left| x - \frac{p}{q} \right| \geq \gamma q^{-2-\mu}$ . Here the choice of the exponent of  $q$  is such that  $\mu$  is always non-negative and can attain the value 0 (e.g. on quadratic irrationals). We denote  $\text{CD}(\gamma, \mu)$  the set of all diophantine  $x$  of exponent  $\mu$  and constant  $\gamma$ .  $\text{CD}(\mu)$  will denote the union  $\cup_{\gamma > 0} \text{CD}(\gamma, \mu)$  and  $\text{CD} = \cup_{\mu \geq 0} \text{CD}(\mu)$ . The complement in  $\mathbb{R} \setminus \mathbb{Q}$  of  $\text{CD}$  is called the set of *Liouville numbers*.

The set of diophantine numbers with a given exponent  $\mu$  has several equivalent descriptions: it is easy to see that

$$\begin{aligned} \text{CD}(\mu) &= \{x \in \mathbb{R} \setminus \mathbb{Q} \mid q_{n+1} = O(q_n^{1+\mu})\} = \{x \in \mathbb{R} \setminus \mathbb{Q} \mid a_{n+1} = O(q_n^\mu)\} \\ &= \{x \in \mathbb{R} \setminus \mathbb{Q} \mid x_n^{-1} = O(b_{n-1}^{-\mu})\} = \{x \in \mathbb{R} \setminus \mathbb{Q} \mid b_n^{-1} = O(b_{n-1}^{-1-\mu})\} \end{aligned}$$

The sets  $\text{CD}(\mu)$  for all  $\mu \geq 0$  are  $\text{SL}(2, \mathbb{Z})$ -invariant and have full Lebesgue measure provided that  $\mu > 0$ .

## References

- [1] Arnold, V.I.: *Proof of a theorem of A. N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations of the Hamiltonian*, Usp. Mat. Nauk, **18**, 13 (1963); Russ. Math. Surv., **18**, 9 (1963).
- [2] Arnold, V.I.: *Small denominators and problems of stability of motion in classical and celestial mechanics*, Usp. Math. Nauk **18** N.6, 91 (1963); Russ. Math. Surv. **18** N.6, 85 (1963).
- [3] Berretti, A. and Gentile, G.: *Scaling properties for the radius of convergence of the Lindstedt series: the standard map*, J. Math. Pures Appl. **78**, 159–176 (1999).
- [4] Berretti, A. and Gentile, G.: *Bryuno function and the standard map*, Comm. Math. Phys. **220**, 623–656 (2001).
- [5] Bruno, A.D.: *Analytical form of differential equations*, Trans. Moscow Math. Soc., **25**, 131–288 (1971); **26**, 199–239 (1972).
- [6] Buff, X. and Chéritat, A.: *The Brjuno function continuously estimates the size of quadratic Siegel disks*, Ann. of Math. **164**, 265–312 (2006).
- [7] 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).
- [8] Davie, A.M.: *The critical function for the semistandard map*, Nonlinearity, **7**, 219–229 (1994)
- [9] Giorgilli, A.: *Quantitative methods in classical perturbation theory*, proceedings of the Nato ASI school “From Newton to chaos: modern techniques for understanding and coping with chaos in N–body dynamical systems”, A.E. Roy e B.D. Steves eds., Plenum Press, New York (1995).
- [10] Giorgilli, A. and Locatelli, U.: *On classical series expansions for quasi-periodic motions*, MPEJ **3** N. 5 (1997).
- [11] 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).
- [12] Gröbner, W.: *Die Lie-Reihen und Ihre Anwendungen*, VEB Deutscher Verlag der Wissenschaften (1967).
- [13] Hardy, G. H. and Wright, E. M.: *An introduction to the theory of numbers*, Oxford University Press (1979).
- [14] Kolmogorov, A.N.: *Preservation of conditionally periodic movements with small change in the Hamilton function*, Dokl. Akad. Nauk SSSR, **98**, 527 (1954). English translation in: Los Alamos Scientific Laboratory translation LA-TR-71-67; reprinted in: Lecture Notes in Physics **93**.
- [15] Khinchin, A.Ya.: *Continued Fractions*, The University of Chicago Press, Chicago-London (1964).
- [16] Marmi, S.: *Critical Functions for Complex Analytic Maps* J. Phys. A: Math. Gen. **23**, 3447–3474 (1990).
- [17] Marmi, S. and Stark, J.: *On the standard map critical function*, Nonlinearity, **5**, 743–761, (1992).

- [18] Marmi, S. and Yoccoz, J.-C.: *Some open problems related to small divisors* in “Dynamical Systems and Small Divisors”, Lecture Notes in Mathematics **1784** 175–191 (2002).
- [19] Marmi, S., Moussa, P. and Yoccoz, J.-C.: *The Brjuno functions and their regularity properties*, Comm. Math. Phys., **186**, 265–293 (1997).
- [20] Moser, J.: *On invariant curves of area-preserving mappings of an annulus*, Nachr. Akad. Wiss. Gött., II Math. Phys. Kl 1962, 1–20 (1962).
- [21] Moser, J.: *Convergent series expansions for quasi-periodic motions*, Math. Ann. **169**, 136–176 (1967).
- [22] Pérez-Marco, R.: *Fixed points and circle maps*, Acta Math. **179** , no. 2, 243–294 (1997).
- [23] Poincaré, H.: *Mémoire sur les courbes définies par une équation différentielle*, Journal de Mathématiques **7**, 375–422 (1881) and **8**, 251–296 (1882).
- [24] Rüssmann, H.: *Über die Iteration analytischer Funktionen*, J. Math. Mech. **17** 523–532 (1967).
- [25] E. Schröder: *Über iterierte Funktionen*, Math. Ann. **3**, 296–322 (1871).
- [26] Siegel, C.L.: *Iterations of analytic functions*, Annals of Math. **43**, 607–612 (1942).
- [27] Siegel, C.L.: *Über die Normalformen analytischer Differentialgleichungen in der Nähe einer Gleichgewichtslösung*, Nachr. Acad. Wiss. Göttingen, Math. Phys. Kl, **IIA** 21–30 (1952).
- [28] Siegel, C.L. and Moser, J.K.: *Lectures in Celestial Mechanics*, Springer–Verlag, Berlin Heidelberg New York (1971).
- [29] Yoccoz, J.-C.: *Théorème de Siegel, nombres de Bruno et polynômes quadratiques* Astérisque **231** 3–88 (1995).
- [30] Yoccoz, J.-C.: *Analytic linearisation of circle diffeomorphisms* in “Dynamical Systems and Small Divisors”, Lecture Notes in Mathematics **1784** 125–173 (2002).