Convergence of the Birkhoff normal form sometimes implies convergence of a normalizing transformation

Rafael de la Llave[*1] and Maria Saprykina[†2]

[1]School of Mathematics, Georgia Institute of Technology
[2]Dept. of Mathematics, KTH Royal Institute of Technoligy

Abstract

Consider an analytic Hamiltonian system near its analytic invariant torus $\mathcal{T}_0$ carrying zero frequency. We assume that the Birkhoff normal form of the Hamiltonian at $\mathcal{T}_0$ is convergent and has a particular form: it is an analytic function of its non-degenerate quadratic part. We prove that in this case there is an analytic canonical transformation—not just a formal power series—bringing the Hamiltonian into its Birkhoff normal form.

1 Introduction

The goal of this paper is to study the convergence of the transformations of an analytic Hamiltonian system in a neighborhood of an invariant torus to the Birkhoff normal form. Here we assume that the frequency vector at the invariant torus is very resonant, hence already at the formal level, the existence of the Birkhoff normal form has obstructions. The main result, Theorem 1 below, will show that if the obstructions for the formal equivalence between the system and its Birkhoff normal form vanish and the normal form is convergent and has a particular form, then the system is analytically equivalent to its normal form. Hence, this result can be considered as a part of the rigidity program: identifying obstructions for a weak form of equivalence whose vanishing implies a stronger form of equivalence.

1.1 Classical theory of normal forms: existence and uniqueness.

Consider an analytic function

$$H(I, \theta) = \langle \lambda_0, I \rangle + \mathcal{O}^2(I), \tag{1.1}$$

where $\theta \in \mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$, $I \in (\mathbb{R}^d, 0)$, $\langle \cdot, \cdot \rangle$ denotes the usual scalar product in $\mathbb{R}^d$, and $\lambda_0 \in \mathbb{R}^d$ is a constant vector called the *frequency vector*. The Hamiltonian system associated to it is $\dot{I} = \partial_\theta H(I, \theta)$, $\dot{\theta} = -\partial_I H(I, \theta)$. Note that we are assuming the standard symplectic form. In particular, the set $\mathcal{T}_0 := \{0\} \times \mathbb{T}^d$ is an invariant torus of this system. We say that $H(I, \theta)$ *has a Birkhoff normal form (BNF)* $N(I)$ in a neighborhood of $\mathcal{T}_0$ if $N(I)$ is a

formal power series, and there exists a *formal* symplectic transformation $\Psi(I, \theta)$, tangent to the identity

$$\Psi(I, \theta) = (I + \mathcal{O}^2(I), \theta + \mathcal{O}(I))$$

such that

$$H \circ \Phi(I, \theta) = N(I)$$

in the sense of formal power series. Any canonical coordinate change $\Phi(I, \theta)$ as above is called a *normalizing transformation*. The following fundamental result is called the Birkhoff normal form [SM71, MHO]. For $H(I, \theta)$ as above, assume that λ_0 satisfies a Diophantine condition: there exist constants (C, τ) such that for all $k \in \mathbb{Z}^d \setminus \{0\}$ we have

$$|\langle \lambda_0, k \rangle| \geq C|k|^{-\tau}. \tag{1.2}$$

Then $H(I, \theta)$ has a (formal) Birkhoff normal form. Moreover, if a normal form exists and λ_0 is rationally independent, then the Birkhoff normal form is unique (up to trivial changes relabelling the actions). Note that the normalizing transformations are not unique, since composing $\Phi(I, \theta)$ with any transformation that preserves I gives a normalizing transformation.

Birkhoff normal form is an important tool in the study of Hamiltonian systems. Already the assumption of existence and nondegeneracy of the normal form has strong dynamical consequences (see, e.g., [EFK15] Th.C). The importance of the BNF becomes even stronger if the normal form is convergent, and even more so if there exists an analytic normalizing transformation.

The standard way of constructing BNF, which we will review in more detail later, is to proceed iteratively, devising transformations that normalize $H(I, \theta)$ up to the coefficients of order I^n. The normalization step involves solving differential equations with analytic conditions. The Diophantine conditions (1.2) can be somewhat weakened to subexponential growth $(\lim_{N \to \infty} \frac{1}{N} \log \sup_{|k| \leq N} |\langle \lambda_0, k \rangle|^{-1} = 0)$.

If λ_0 is resonant, one cannot guarantee the existence of the Birkhoff normal form even at the level of formal power series, since there may be some terms in the formal power series of H that cannot be eliminated by a canonical transformation. On the other hand, there are, of course, systems (e.g the BNF itself, or changes of variables from it) for which one can construct a BNF even in the resonant case. Then one speaks of the Birkhoff-Gustavson normal form [Gu66].

Analogous definitions and statements hold true for symplectic maps in a neighborhood of a fixed point. Even if the formal elimination procedures are very similar, the analysis is very different. Handy references for the classical theory of Birkhoff normal forms are [SM71, MHO, Mu, EFK13, EFK15].

1.2 Generic divergence both of the Birkhoff Normal Form and the normalizing transformation.

The BNF and the normalizing transformations are constructed as formal power series. The following natural questions are of great importance: the first one is whether the BNF converges for Hamiltonians in a certain class. The second—whether there is a convergent normalizing transformation.

Concerning the first question, R. Perez-Marco [PM] proved the following dichotomy: for any given nonresonant quadratic part, either the BNF is generically divergent or it always converges. The original proof was done in the setting of Hamiltonian systems having a non resonant elliptic fixed point. The extension of this result to the case of the torus, that is not completely straightforward, has been worked out by R. Krikorian, see Theorem 1.1 in [Kri].

Up to very recently it was unclear which of the possibilities is actually realized. A large progress has been made by R. Krikorian [Kri], who proved that there exists a real analytic symplectic diffeomorphism f of a two-dimensional annulus such $f(\mathbb{T} \times \{0\}) = (\mathbb{T} \times \{0\})$, $f(\theta, 0) = (\theta + \omega_0, 0)$ with ω_0 Diophantine and having a non-degenerate *divergent* Birkhoff normal form. Combined with the aforementioned result of Perez-Marco, this implies that Birkhoff Normal Form of an analytic Hamiltonian is "in general" divergent.

Concerning the normalizing transformations, H. Poincaré proved that they are divergent for a generic Hamiltonian. C. L. Siegel proved the same statement in a neighborhood of an elliptic fixed point (in fact, for a larger class of Hamiltonians than just generic, [Si54]). This is implied by showing that the orbit structure of the map in any neighborhood is very different from that of the Birkhoff normal form (which is integrable). Analogous results for symplectic maps near an elliptic fixed point appear in [Rü59]. Very different arguments showing divergence of normalizing transformations for generic systems appear in [Ze73] and for some concrete polynomial mappings in [Mo60].

1.3 Convergence of the transformations under the Diophantine conditions for some particularly simple BNF

There are classes of Hamiltonians for which we can guarantee the convergence of the normalizing transformation. The following influential rigidity result was proved independently by A. D. Bruno [Br71] and H. Rüssmann [Rü67]. Note that the main assumption is that the (in principle only formal) BNF is of a particular kind.

Consider an analytic Hamiltonian $H(I, \theta)$, whose frequency λ_0 satisfies a Diophantine condition (1.2). Assume moreover that the Birkhoff normal form $N(I)$ of $H(I, \theta)$ is a *formal* function B of one single variable $\Lambda_0 := \langle \lambda, I \rangle$, i.e.,

$$N(I) = B(\Lambda_0(I)).$$

Then there exists an *analytic* normalizing transformation, and the BNF is, in fact, analytic.

We remark that Bruno proves the above result under a weaker condition on λ_0 than (1.2). For analogous statements in the case of invariant tori see [Br89]. Other modifications can be found in [Rü02, Rü04]. This result has been recently generalised to a much more general context by Eliasson, Fayad and Krikorian [EFK13, EFK15]. We stress that in all these works mentioned above, λ_0 is assumed to be non-zero and the crucial assumption is that λ_0 satisfies a Diophantine-type condition and that the BNF is of a very simple form.

1.4 "Sometimes" convergence of the BNF implies convergence of a normalizing transformation.

Our main result is close in spirit to the above works, but it *does not rely on a Diophantine condition*. In fact, we consider a special class of diffeomorphisms such that the frequency λ_0 is zero. Thus, the BNF is degenerate in the previous sense. But within this class of Hamiltonians we just use a standard non-degeneracy assumption on the quadratic part. Namely, we prove the following.

Theorem 1. *Assume the following:*

(A_1) $H(I,\theta)$ *has a formal Birkhoff normal form* $N(I)$ *that starts with quadratic terms in* I, *i.e there exists a formal symplectic change of variables* $\Psi(I,\theta)$, *tangent to the identity, i.e.* $\Psi(I,\theta) = (I + \mathcal{O}^2(I), \phi + \mathcal{O}(I))$, *such that*

$$H \circ \Psi(I,\theta) = N(I) = N_0(I) + \mathcal{O}^3(I)$$

in the sense of power series.

(A_2) $N_0(I) = I^{tr}\Omega I$ *(for some symmetric* Ω*) is non-degenerate:* $\det \Omega \neq 0$.

(A_3) $N(I) = B(N_0(I)) = N_0 + \sum_{j=2}^{\infty} b_j (N_0(I))^j$ *where* B *is an analytic function.*

Then there exists an invertible analytic symplectic transformation

$$\Phi(I,\theta) = (I + \mathcal{O}^2(I), \phi + \mathcal{O}(I))$$

such that

$$H \circ \Phi(I,\theta) = N(I). \tag{1.3}$$

Note that we start from a resonant torus, so that the existence of a BNF of the form we assume, requires vanishing of (formal) obstructions. Hence, our main result can be reformulated as saying that the formal assumptions imply convergence of the normalizing transformation.

Similar rigidity statements have appeared in other contexts. In [Po92, Ch. 5], H.Poincaré studied the formal power series of canonical transformations, which send a family of Hamiltonian systems into a family of integrable systems (in the sense of power series). In [Po92] it was shown that these formal power series do not exist unless there are some conditions (which are not met in the three body problem for arbitrary masses). The non-existence of formal power series, a fortiori implies the non-existence of analytic families of analytic transformations integrating the three body problem.

The paper [Ll] proved a converse to the result in [Po92]: if the system satisfies a very specific and generic non-degeneracy condition, then, existence of a formal power series that integrates the family of transformations in the sense of power series implies existence of a convergent one.

Assumption A_3 is there for technical purposes, see Sec. 3.3. Note that it is trivial for $d = 1$. This assumption reminds of that of Rüssmann in [Rü67, Rü02, Rü04].

The assumption that the Birkhoff normal form is a function of N_0 has been discussed in [Ga] under the name of *relative integrability*. Two Hamiltonian dynamical systems are

relatively integrable when one of them can be obtained from the other by a symplectic change of coordinates and a reparameterization of the time which only depends on the total energy. That is, the orbit structures of the two systems in an energy surface are equivalent up to a change of scale of time. The paper [Ga] includes several arguments for why the notion of relative integrability is natural when discussing formal equivalence. In the present paper, however, the focus lies on the notion of equivalence under a symplectic change of variables. We show that, for a certain class of systems, equivalence in the sense of formal power series implies equivalence in the sense of analytic canonical changes of variables. Hence, our main result can be understood as a rigidity result. The class of systems for which this rigidity result holds can be succinctly described as the set of systems that relatively integrable with respect to the main term.

In the context of formal equivalence implying analytically convergent equivalence, it is natural to formulate:

Conjecture. *Assume that an analytic Hamiltonian $H(I, \theta)$ as in (1.1) has a convergent BNF that satisfies the non-degeneracy assumption that the frequency map is a local diffeomormphism. Then there is a convergent normalizing transformation.*

Note that the problems studied in [Rü67] and [Br71] do not satisfy the hypothesis of the conjecture, even though they satisfy the conclusions.

In the other direction one can construct examples [S] of analytic maps near a hyperbolic fixed point such that the Birkhoff normal form is quadratic (in the above notations, $N = \Lambda_0$) with a non-resonant set of eigenvalues, and any normalizing transformation to the normal form diverges. In these examples, the eigenvalues form carefully chosen Liouville vectors. That is, the paper [S] shows that, depending on the Diophantine conditions, quadratic normal forms may be rigid or not. The models in [S] do not satisfy the hypothesis of the conjecture above.

1.5   Overview of the proof.

The standard method of obtaining the Birkhoff Normal form is an iterative procedure in which we construct the transformations order by order: at the n-th step of the procedure one computes the n-th order terms in the Taylor expansions, assuming that all the terms of lower orders are computed. It would appear natural to follow this scheme and try to *estimate* the transformations at each step of the recursive procedure. Unfortunately, this seems technically unfeasible. One of the main complications in any possible proof of convergence of the transformations is that even if the BNF is unique, the formal transformations Φ_N are very far from unique (Since the BNF depeds only on the actions, the Φ_N can be composed with any canonical transformation which moves the angles but preserves the actions. So, an essential ingredient of any proof of convergence should be a especification of how to choose the normalizing transformations.

In this paper we use a quadratically convergent method in which we double the number of known coefficients at each step. Roughly – see more details in the next paragraphs – we will show that if the formal obstructions vanish we can choose a sequence of canonical transformations that proceed to converge quadratically: doubling the order of the BNF at every step of the construction. More importantly, there is a specific choice of the transformation that satisfies very explicit bounds. The bounds on the new transformation

in terms of the remainder turn out to involve a loss of derivatives. Therefore we need to implement a Nash-Moser scheme to estimating the important objects in a sequence of domains which decrease slowly.

Here is a short overview of the proof; all the necessary notations are introduced in the next section. At the n-th step of the iterative procedure we will start with a Hamiltonian of the form

$$H_n(I, \theta) = N_n(I) + \widetilde{R_n}(I, \theta),$$

where $N_n(I)$ is a polynomial in I of degree $m_n = 2^n + 1$, and the remainder term $\widetilde{R_n}$ is small in the following sense: for a certain domain-dependent norm, introduced in Sec. 2.1.1, for a certain small δ_n (we assume $\delta_n \to 0$ with $n \to \infty$) and $\kappa > 0$ the remainder term satisfies $|\widetilde{R_n}|_{\rho_n, \rho_n} \leq \delta_n^\kappa$.

At this step we construct a symplectic change of coordinates Φ_n, such that

$$H_n \circ \Phi_n(I, \theta) = N_{n+1}(I) + \widetilde{R_{n+1}}(I, \theta),$$

where N_{n+1} has degree $m_{n+1} = 2m_n - 1$, and $|\widetilde{R_{n+1}}|_{\rho_{n+1}, \rho_{n+1}} \leq \delta_{n+1}^\kappa = 2^{-\kappa} \delta_n^\kappa$.

We construct Φ_n as a time one map of a the flow of a Hamiltonian vector field F_n. The main ingredient consists in constructing and estimating the norm of F_n (and thus Φ_n), which is found as a solution of a certain homological equation (see (3.1) and in a simplified form (4.1)). In general, this equation may not have even a formal solution unless some constraints are met. However, the assumption of Theorem 1 implies that this equation does have a formal solution. The key observation in this paper is the following: if this homological equation has a *formal solution*, then it also has an *analytic solution with tame estimates* for it (in the sense of Nash-Moser theory). This statement is the contents of Lemma 6. We note that the tame estimates use an argument different from the matching of powers.

The procedure can be repeated, because the main assumption used to show the existence of solutions of the Newton equation is that there is a formal solution to all orders. This assumption is clearly preserved if we make any analytic change of variables. Once we know that the Newton procedure can be repeated infinitely often, the convergence is more or less standard.

2 Notations and a step of induction.

2.1 Notations.

2.1.1 Norms and majorants.

Let $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$ be a d-dimensional torus, and for $\sigma > 0$ consider its complex extension $\mathbb{T}_\sigma^d = \left(\mathbb{R}^d + (-\sigma, \sigma) \sqrt{-1} \right) / \mathbb{Z}^d$. Let $\mathbb{D}_\rho^d = \{ I \in \mathbb{C}^d : |I| < \rho \}$ be a complex disc, and define the "d-dimensional annulus"

$$\mathbb{A}_{\rho, \sigma} := \mathbb{D}_\rho^d \times \mathbb{T}_\sigma^d.$$

Let $\mathcal{O}(\mathbb{A}_{\rho, \sigma})$ be the set of functions holomorphic in $\mathbb{A}_{\rho, \sigma}$ that are real symmetric, i.e., such that $\overline{f(\bar{I}, \bar{\theta})} = f(I, \theta)$ (where the bar stands for the complex conjugate). We use supremum norms over $\mathbb{A}_{\rho, \sigma}$, denoted by $\|f\|_{\rho, \sigma}$. In the same way we define the set $\mathcal{O}(\mathbb{D}_\rho)$ with the corresponding norm $\|f\|_\rho$ being the sup-norms over the disc $\mathbb{D}_\rho^d$.

For a function $f \in \mathcal{O}(\mathbb{A}_{\rho,\sigma})$ consider its Taylor-Fourier representation in the powers of I: $f(I,\theta) = \sum_{j \in \mathbb{N}^d} \sum_{k \in \mathbb{Z}^d} f_{j,k} e^{2\pi i \langle k,\theta \rangle} I^j$. Consider a majorant for f of the form

$$\widehat{f}(I) = \sum_{j \in \mathbb{N}^d} \sum_{k \in \mathbb{Z}^d} |f_{j,k}| I^j e^{2\pi |k| \sigma}$$

We denote by $|f|_{\rho,\sigma}$ the norm of the corresponding majorant $\widehat{f}(I)$:

$$|f|_{\rho,\sigma} = \|\widehat{f}\|_{\rho,\sigma}.$$

Clearly, $\|f\|_{\rho,\sigma} \le |f|_{\rho,\sigma}$. Analogous notation $|f|_\rho$ corresponds to the norm $\|f\|_\rho$ above. In what follows we will mostly have $\sigma = \rho$.

2.1.2 Important constants for the iterative procedure.

□ Let $\rho_0 = \min\{1, \rho\}$,

□ The order of polynomials involved in the n-th step of the iterative procedure is

$$m_n = 2^n + 1.$$

□ The norm of the rest term $\widetilde{R_n}$ at the n-th step will be estimated as $|\widetilde{R_n}|_{\rho_n} \le \delta_n^\kappa$. Let

$$\kappa = d + 6,$$
$$b = 2^{-(\kappa+3)}$$
$$\delta_0 = \rho_0 b 2^{-3} = \rho_0 2^{-(\kappa+6)}$$
$$\delta_{n+1} = 2^{-1} \delta_n.$$

□ Finally, let

$$q_n = (2b)^{2^{-(n+1)}},$$

and

$$\rho_{n+1} = (\rho_n - 3\delta_n) q_n.$$

2.1.3 Polynomials.

In the iterative procedure we will work with polynomials in I whose coefficients depend on θ.

□ Let

$$N_0(I) = I^{tr} \Omega I \tag{2.1}$$

where Ω is a symmetric non-degenerate matrix: $\det \Omega \ne 0$.

□ An expression $M = f(\theta) I^k$ (where k is a multi-index) is called a monomial.

□ We will say that a monomial $M_{k,l} = I^k e^{2\pi i \langle l,\theta \rangle}$ is *resonant* if it satisfies $\{N_0, M\} = 0$.

□ $R^{[j]}(I,\theta)$ stands for a homogeneous polynomial in I of degree j with coefficients depending on θ:

$$R^{[j]}(I,\theta) = \sum_{|k|=j} r_k(\theta) I^k.$$

7

□ We also use notation $R^{[m,n]}$ to denote the range of degrees in I:

$$R^{[m,n]}(I,\theta) = \sum_{j=m}^{n} R^{[j]}(I,\theta), \quad R^{[\geq m]}(I,\theta) = \sum_{j=m}^{\infty} R^{[j]}(I,\theta).$$

Let m_n be as above. The following functions will be of special importance.

□ The normal form $N(I)$ is assumed to have the form

$$N(I) = B(N_0(I)) = N_0(I) + \sum_{j=2}^{\infty} b_j(N_0(I))^j. \qquad (2.2)$$

Denote

$$N_n = N^{[2,m_n]} = (B(N_0))^{[2,m_n]}; \qquad (2.3)$$

in particular, since $m_0 = 2$, $N_0 = N_0^{[2,m_0]} = N_0^{[2]}$ is quadratic.

□ The rest term at the n-th inductive step is $\widetilde{R_n}(I,\theta)$:

$$\widetilde{R_n} = \widetilde{R_n}^{[>m_n]}. \qquad (2.4)$$

□ We will also need polynomials in I with θ-dependent coefficients: $R_n(I,\theta)$ and $F_n(I,\theta)$ of the following degrees:

$$R_n = R_n^{[m_n+1,m_{n+1}]}, \quad F_n = F_n^{[m_n,m_{n+1}-1]}. \qquad (2.5)$$

2.2 Base of induction: an equivalent problem.

Lemma 1. *Suppose that*

$$H(I,\theta) = N_0(I) + \widetilde{R_0}(I,\theta) \in \mathcal{O}(\mathbb{A}_{\rho,\sigma}),$$

where $|\widetilde{R_0}|_{\rho,\sigma} \leq \delta$, and there exists a formal (resp., analytic) symplectic transformation

$$\Psi(I,\theta) = (\phi(I,\theta), \psi(I,\theta)) = (I + \mathcal{O}^2(I), \theta + \mathcal{O}(I))$$

such that

$$H \circ \Psi(I,\theta) = N(I) = N_0(I) + \sum_{j=2}^{\infty} b_j(N_0(I))^j.$$

Then for any $a > 0$ there exists a Hamiltonian $\widehat{H}(I,\theta)$ and a formal (resp., analytic) symplectic transformation $\widehat{\Psi}(I,\theta) = (I + \mathcal{O}^2(I), \theta + \mathcal{O}(I))$ such that

$$\widehat{H} \circ \widehat{\Psi}(I,\theta) = N_0(I) + \widehat{R_0}(I,\theta) \in \mathcal{O}(\mathbb{A}_{\frac{1}{a}\rho,\sigma}),$$

where $|\widehat{R_0}|_{\frac{1}{a}\rho,\sigma} \leq a\delta$, and

$$N(I) = N_0(I) + \sum_{j=2}^{\infty} b_j a^{2(j-1)}(N_0(I))^j.$$

Proof. Define $\widehat{H}(I,\theta) = \frac{1}{a^2}H(aI,\theta)$, and $\widehat{\Psi}(I,\theta) = \left(\frac{1}{a}\phi(aI,\theta), \psi(aI,\theta)\right)$. It can be verified directly that $\widehat{\Psi}$ is symplectic and tangent to the identity. Moreover,

$$\widehat{H} \circ \widehat{\Psi}(I,\theta) = \frac{1}{a^2}H(\phi(aI,\theta), \psi(aI,\theta)) = N_0(I) + \sum_{j=2}^{\infty} b_j a^{2(j-1)}(N_0(I))^j.$$

$\square$

2.3 Induction step.

While the base of induction is given by formula (2.12), the step of the iterative procedure is provided by the following proposition.

Proposition 1. *For a fixed $n > 0$, let m_n, ρ_n and δ_n be as in Sec. 2.1.2 above. Suppose that $H_n(I,\theta)$ is formally conjugated to the BNF of the form (2.2):*

$$N(I) = N_0(I) + \sum_{j=2}^{\infty} b_j(N_0(I))^j,$$

and the normal form satisfies:

$$|N^{[m_n+j]}|_{\rho_n} < \delta_n^{\kappa+1}, \quad j = 0, \ldots m_n, \tag{2.6}$$

and denoting $g_{2j}(I) = jb_j(N_0(I))^{j-1}$, we assume

$$|g_j|_{\rho_n} \leq \frac{1}{4^j}, \quad j = 1, \ldots, m_n; \tag{2.7}$$

Suppose that

$$H_n(I,\theta) = N_n(I) + \widetilde{R_n}(I,\theta),$$

where $N_n(I) = (B(N_0(I)))^{[2,m_n]}$ and $\widetilde{R_n} = \widetilde{R_n}^{[>m_n]}$ satisfies

$$|\widetilde{R_n}|_{\rho_n,\rho_n} \leq \delta_n^{\kappa}.$$

Then there exists a symplectic change of coordinates $\Phi_n : (I',\theta') \mapsto (I,\theta)$,

$$\Phi_n(I',\theta') = (U^{(n)}(I',\theta'), V^{(n)}(I',\theta')),$$

given by a Hamiltonian $F_n = F_n^{[m_n, m_{n+1}-1]}$ such that

$$H_{n+1}(I',\theta') := H_n \circ \Phi_n(I',\theta') = N_{n+1}(I') + \widetilde{R_{n+1}}(I',\theta'), \tag{2.8}$$

where $N_{n+1}(I') = N^{[2,m_{n+1}]}(I')$, $\widetilde{R_{n+1}}(I',\theta') = \widetilde{R_{n+1}}^{[>m_{n+1}]}(I',\theta')$, and

$$|\widetilde{R_{n+1}}|_{\rho_{n+1},\rho_{n+1}} \leq \delta_{n+1}^{\kappa}. \tag{2.9}$$

Moreover, $\Phi_n(I',\theta') = (U^{(n)}(I',\theta'), V^{(n)}(I',\theta'))$ satisfies

$$\sum_{j=1}^{d} \|U_j^{(n)}(I',\theta') - I_j'\|_{\rho_n-3\delta_n,\rho_n-3\delta_n} + \|V_j^{(n)}(I',\theta') - \theta_j'\|_{\rho_n-3\delta_n,\rho_n-3\delta_n} < \delta_n, \tag{2.10}$$

9

and the inverse map, $\Phi_n^{-1}(I,\theta) := (U^{(-n)}(I,\theta), V^{(-n)}(I,\theta))$, satisfies

$$\sum_{j=1}^{d} \|U_j^{(-n)}(I,\theta) - I_j\|_{\rho_n - 3\delta_n, \rho_n - 3\delta_n} + \|V_j^{(-n)}(I,\theta) - \theta_j\|_{\rho_n - 3\delta_n, \rho_n - 3\delta_n} < \delta_n. \qquad (2.11)$$

The proof of this proposition constitutes the main technical tool of this paper. It implies Theorem 1 in a standard way. See, e.g., [Rü67], pp 61-63). For convenience, we give a proof below.

2.4 Proof of Theorem 1.

Lemma 1 permits us to assume without loss of generality that for the given Hamiltonian $H_0(I,\theta) := H(I,\theta) = N_0(I) + \widetilde{R_0}(I,\theta)$,

$$|\widetilde{R_0}|_{\rho_0, \rho_0} \leq \delta_0^{\kappa}. \qquad (2.12)$$

Since the function B is analytic, the same lemma permits us to assume that (2.6) and (2.7) hold for each n.

The step of induction is provided by Proposition 1. Since H_n is formally reducible to the normal form N, the same can be said about H_{n+1}.

Repetition of this process leads to a sequence of transformations

$$T_n = \Phi_0 \circ \Phi_1 \circ \cdots \circ \Phi_{n-1}.$$

Let us show that T_n converges to the desired coordinate change $\Phi = T_\infty$, analytic in the polydisc $\mathbb{A}_{\rho_\infty, \rho_\infty}$, where $\rho_0 b < \rho_\infty < \rho_0$. Indeed, with the notations of Sec. 2.1.2,

$$3\sum_{k=0}^{\infty} \delta_k \leq 3 \cdot 2\delta_0 < 3 \cdot 2\rho_0 b 2^{-3} < \rho_0 b.$$

Then for any n we have

$$\rho_{n+1} = q_n(\rho_n - 3\delta) \geq \rho_0 \prod_{j=0}^{n} q_j - 3\sum_{j=0}^{n} \delta_n \geq \rho_0 \prod_{j=0}^{\infty} q_j - 3\sum_{j=0}^{\infty} \delta_n \geq \rho_0 2b - 3 \cdot 2\delta_0 > b\rho_0$$

It is left to prove that T_n converges of to an analytic function T_∞, satisfying (1.3). Denote the variables, involved in the n-th step of the induction by $w_{n-1} = (I,\theta)$ and $w_n = (I', \theta')$, where

$$w_n = \Phi_{n-1}^{-1} w_{n-1}.$$

In these notations,

$$w_0 = \Phi_0 \circ \Phi_1 \circ \cdots \circ \Phi_{n-1} w_n = T_n w_n.$$

Now, for $w_n = (I', \theta')$ we have

$$H \circ T_n(I', \theta') = N_n(I') + \widetilde{R_n}(I', \theta').$$

Since $(\Phi_n(I', \theta') - (I', \theta'))$ starts with the terms of degree 2^n in I', for each j the expansion of $(T_n(I', \theta') - T_{n+j}(I', \theta'))$ starts with the terms of degree 2^n in I'. This implies that the

10

sequence of maps T_n formally converges, when $n \to \infty$, to a formal map T_∞ such that (1.3) holds:

$$H \circ T_\infty(I', \theta') = N(I').$$

We still need to show that T_∞ is analytic. It is more convenient to prove that the maps

$$T_n^{-1} := \Phi_{n-1}^{-1} \circ \cdots \circ \Phi_1^{-1} \circ \Phi_0^{-1}$$

converge to an analytic map T_∞^{-1}.

By Proposition 1, the map

$$w_{n+1} = \Phi_n^{-1} w_n$$

is analytic in $\mathbb{A}_{\rho_0 b/2, \rho_0 b/2}$, and for all n we have:

$$|\Phi_n^{-1} w_n - w_n|_{\rho_0 b/2, \rho_0 b/2} \leq \delta_n,$$

since $\rho_n - 3\delta \geq \rho_{n+1} > \rho_0 b$ for all n. Therefore, the map T_n^{-1} such that

$$w_n = T_n^{-1} w_0$$

is analytic in $\mathbb{A}_{\rho_0 b/4, \rho_0 b/4}$, and for such w_0 we have

$$|T_n^{-1} w_0| \leq \sum_{j=0}^{n-1} |T_j^{-1}(w_j) - w_j| + |w_0| \leq \sum_{j=0}^{\infty} \delta_j + \rho_0 b/4 \leq \rho_0 b/2.$$

Estimate

$$|T_{n+m}^{-1}(w_0) - T_n^{-1}(w_0)|_{\rho_0 b/4, \rho_0 b/4} \leq \sum_{j=n}^{n+m-1} |T_j^{-1}(w_j) - w_j|_{\rho_0 b/4, \rho_0 b/4} \leq \sum_{j=n}^{\infty} \delta_j = 2^{1-n} \delta_0$$

implies the convergence of the sequence of maps T_n^{-1} to an analytic map T_∞^{-1} in $\mathbb{A}_{\rho_0 b/4, \rho_0 b/4}$. Since the formal inverse of T_∞^{-1} is the series T_∞, the latter also defines an analytic function, providing the desired coordinate change. We set $\Phi = T_\infty$ in the notations of Theorem 1. $\square$

3   Formal analysis.

Here we start the proof of Proposition 1 by the formal analysis of the iterative procedure.

3.1   Iterative Procedure.

Given H_n as in Proposition 1, we will construct Φ_n as the time one map of the flow of a Hamiltonian F_n, i.e., $\Phi_n = X_{F_n}^1$ where $X_{F_n}^t$ is the flow defined by

$$\dot{I} = F_\theta(I, \theta), \quad \dot{\theta} = -F_I(I, \theta).$$

In this case, Φ_n is automatically symplectic.

Notice that the normalising transformation Φ_n, as well as the corresponding generating function F_n, is not unique (one can compose with rotations in the angles which preserve

11

the actions, for example). Clearly, the transformation that converges has to be very carefully chosen.

In the following Lemma 2 we show that if a (formal) normalizing transformation exists, then there exists (another) normalizing transformation of a special kind. Namely, such that the corresponding generating function is a polynomial (in the sense of section 2.1.3), $F_n = F_n^{[m_n, m_{n+1}-1]}$ and free from resonant monomials (see notations in Sec. 2.1.3).

The idea of the proof is that we can always move the formal normalizing transformation by composing with some transformations that do not change the normal form. Therefore, we can ensure that the normalizing transformations belong to a space which is transversal to the space spanned by resonant monomials. Note that in the proof of Lemma 2 we use crucially the fact that the normal form is a function of N_0 so that the resonant terms are the same at all orders.

There are some analogies between Lemma 2 and Proposition 2.6 in [Ll], but that result is significantly less delicate since there is an extra parameter that controls the smallness. In our case, the variable I controls both the smallness and the distance to the origin at the same time.

Let $\{\cdot, \cdot\}$ denote the standard Poisson bracket. Recall that for a differentiable function G it holds:

$$\frac{d}{dt} G \circ X_F^t = \{G, F\} \circ X_F^t.$$

Lemma 2. *Suppose that for $H(I, \theta)$ there exist $N_{2m}(I) = N_0 + B(N_0)$ with $B(X) = \sum_{j=2}^{m} b_j X^j$, $R(I, \theta) = R^{[>2m]}(I, \theta)$ and $G(I, \theta) = \mathcal{O}^2(I)$ such that $\Psi := X_G^1$ satisfies*

$$H \circ \Psi(I, \theta) = N_{2m}(I) + R(I, \theta).$$

1. *Then there exists $\tilde{G}(I, \theta)$, that is free from resonant monomials of order $< 2m$, such that $\tilde{\Psi} := X_{\tilde{G}}^1$ normalises H to the same normal form, i.e., for some $\tilde{R}(I, \theta) = (\tilde{R})^{[>2m]}(I, \theta)$ we have:*

$$H \circ \tilde{\Psi}(I, \theta) = N_{2m}(I) + \tilde{R}(I, \theta).$$

2. *If, an addition to the previous assumption, we have that the original $H(I, \theta)$ has the form*

$$H(I, \theta) = N_m(I) + R^{[>m]}(I, \theta),$$

   where $N_m = N_m^{[2, \dots, m]}$, then there exists a polynomial $F = F^{[m, 2m-2]}$, that is free from resonant monomials, such that $\Phi := X_F^1$ normalises H to the same normal form, i.e., for some $\tilde{\tilde{R}}(I, \theta) = \tilde{\tilde{R}}^{[>2m]}(I, \theta)$ we have:

$$H \circ \Phi(I, \theta) = N_{2m}(I) + \tilde{\tilde{R}}(I, \theta).$$

Proof: (1). All the calculations below are made in the sense of formal Taylor-Fourier expressions. Suppose that $K(I, \theta)$ is such that $\{N_0, K\} = 0$. Notice that in this case

$\{N_{2m}, K\} = B'(N_0)\{N_0, K\} = 0$. Use $K(I, \theta)$ as a Hamiltonian to define $k(I, \theta) := X_K^1$. Then by Taylor formula we have:

$$H \circ \Psi \circ k = (N_{2m} + R) \circ k = (N_{2m} + R) \circ X_K^t \Big|_{t=1} = N_{2m} + R + \{(N_{2m} + R), K\}$$

$$+ \frac{1}{2}\{\{(N_{2m} + R), K\}, K\} + \cdots = N_{2m} + R_1,$$

where $R_1(I, \theta) = R_1^{[>2m]}(I, \theta)$.

It is a classical fact that the composition $\Psi \circ k$ in the sense of formal power series is the time-one map of another Hamiltonian given by the Cambell-Baker-Dynkin formula [Dragt, Appendix C], [LlMM, Appendix]; here we denote it by CBD formula. Note that in these references the usual notation for the Hamiltonian vector field defined by G is $\mathcal{L}_G$, and $\exp(\mathcal{L}_G)$ stands for its time one map. In the present paper the same map is denoted by X_G^1. Now, suppose that $\Psi = X_G^1$ and $k = X_K^1$. CBD formula implies that the composition of these maps satisfies:

$$\tilde{\Psi} := \Psi \circ k = X_{\tilde{G}}^1, \quad \text{where}$$

$$\tilde{G} = G + K + \frac{1}{2}\{G, K\} + \frac{1}{12}\{G, \{G, K\}\} - \frac{1}{12}\{K, \{K, G\}\} + \cdots$$

The last sum is to be understood in the sense of formal power series in I.

To prove Lemma 2, we use CBD formula, and choose K recursively (order by order in I) so that $\tilde{G}$ has no resonant terms up to order $2m$. At each step of the recursion we choose $(-K(I, \theta))$ to be equal to the lowest order resonant term of G, and set $\tilde{G}$ to be the new G. As we saw above, the map $\tilde{\Psi} = \Psi \circ K$, used as a normalization map, brings H to the same normal form as Ψ did. But its generating Hamiltonian $\tilde{G}$ has no lower order resonant monomials. Iterating this procedure, we get a normalization with the desired property.

(2). Since we can normalise $H = N_m + R^{[>m]}$ to N_{2m} with the help of the generating function $G = \mathcal{O}^2(I)$, then, by *(1)*, we can also achieve the normalization using the transformation $\tilde{\Psi}$ generated by a resonance-free Hamiltonian $\tilde{G}$. Note that $\tilde{G} = \mathcal{O}^2(I)$.

By the Taylor formula for power series, we have:

$$H \circ \tilde{\Psi} = (N_m + R^{[>m]}) \circ \tilde{\Psi} = (N_m + R^{[>m]}) \circ X_{\tilde{G}}^t \Big|_{t=1} = N_m + R^{[>m]}$$

$$+ \{(N_m + R^{[>m]}), \tilde{G}\} + \frac{1}{2}\{\{(N_m + R^{[>m]}), \tilde{G}\}, \tilde{G}\} + \cdots = N_{2m} + R_1.$$

Since $\tilde{G}$ is resonance-free, any monomial P in $\tilde{G}$ gives a non-zero impact $\{N_0, P\}$ to the sum above, whose order in I is strictly larger than the order of P. By comparing the orders of the coefficients in I we see that the lowest possible order of a monomial in $\{N_0, \tilde{G}\}$ is the same as that in $R^{[>m]}$, and hence $\tilde{G} = \tilde{G}^{[\geq m]}$. Finally notice that the reduced generating function $F := \tilde{G}^{[m, 2m-2]}$ produces the same normal form.

$\square$

The following lemma introduces the notations used in the proof of the Main Theorem. Here we use the results of Lemma 2 to relate the conjugating function to the solutions of the homological equation (3.1) below.

Lemma 3. *Adopt the notations for the degrees of polynomials from Sec. 2.1.3 (in particular, $N_n = N^{[2,m_n]}$ as in 2.3, and $R_n = R_n^{[m_n+1,m_{n+1}]}$). Let $B(X) = \sum_{j=1}^{\infty} b_j X^j$. Suppose that H_n has the form*

$$H_n = N_n + \widetilde{R_n} = N_n + R_n + \widetilde{R_n}^{[>m_{n+1}]}.$$

where $N_n(I) = N_0 + B(N_0)^{[4,m_n]}$.

Suppose that there exists $G(I,\theta) = \mathcal{O}^2(I)$ such that $\Psi := X_G^1$ satisfies

$$H \circ \Psi(I,\theta) = N_{m+1}(I) + R(I,\theta).$$

Then there exists a polynomial (in I) $F_n = F_n^{[m_n,m_{n+1}-1]}$ with the following properties: the time one map $\Phi_n := X_{F_n}^1$ satisfies

$$H_{n+1} := H_n \circ \Phi_n = N_{n+1} + \widetilde{R_{n+1}},$$

and F_n satisfies

$$\{N_n, F_n\}^{[m_n+1,m_{n+1}]} + R_n + N_n - N_{n+1} = 0, \tag{3.1}$$

and

$$\widetilde{R_{n+1}} := A_n + B_n + C_n,$$

where

$$A_n := \widetilde{R_n}^{[>m_{n+1}]} \circ \Phi_n, \quad B_n := \int_0^1 \{(1-t)\{N_n, F_n\} + R_n, F_n\} \circ X_{F_n}^t dt, \tag{3.2}$$

$$C_n = (\{N_n, F_n\})^{[>m_{n+1}]}. \tag{3.3}$$

Notice that the expressions for A_n, B_n, C_n start with terms of order $m_{n+1}+1$, and hence, $\widetilde{R_{n+1}} = \widetilde{R_{n+1}}^{[>m_{n+1}]}$, as needed.

Proof. Let $m = m_n = 2^n + 1$. Then $m_{n+1} = 2m - 1$. With the notations for the degrees of polynomials from Sec. 2.1.3, Lemma 2 implies that there exists a polynomial $F_n = F_n^{[m_n,m_{n+1}-1]}$ such that $\Phi_n := X_{F_n}^1$ satisfies $H_n \circ \Phi_n = N_{n+1} + \widetilde{R_{n+1}}$. By the Taylor formula we have:

$$
\begin{aligned}
H_n \circ \Phi_n =&(N_n + R_n + \widetilde{R_n}^{[>m_{n+1}]}) \circ X_{F_n}^t \Big|_{t=1} = N_n + \{N_n, F_n\} + R_n + \\
&\int_0^1 \{(1-t)\{N_n, F_n\} + R_n, F_n\} \circ X_{F_n}^t \, dt + \widetilde{R_n}^{[>m_{n+1}]} \circ \Phi_n \\
=&N_{n+1} + \widetilde{R_{n+1}}.
\end{aligned}
\tag{3.4}
$$

Notice that by extracting all the terms of orders $m_n + 1, \ldots, m_{n+1}$ from the equation above, one gets the cohomological equation (3.1).

$\square$

3.2 Homological equation order by order.

Here we rewrite equation (3.1) as a (finite) set of equations for each degree of I. Equations corresponding to degrees $m_n + 1, \ldots, m_{n+1}$ will formally determine F_n (they are written out explicitly in (3.5)). The rest of equations define C_n (which is a part of the new remainder term). Equating coefficients with the same homogeneous degree in I in both sides of (3.4) we obtain for the degrees from $m_n + 1$ to m_{n+1} the following recursive formula (we write m instead of m_n for typographic reasons):

$$
\begin{aligned}
&\{N_0, F^{[m]}\} + R^{[m+1]} = N^{[m+1]}, \\
&\{N_0, F^{[m+1]}\} + \{N^{[3]}, F^{[m]}\} + R^{[m+2]} = N^{[m+2]}, \\
&\{N_0, F^{[m+2]}\} + \{N^{[4]}, F^{[m]}\} + \{N^{[3]}, F^{[m+1]}\} + R^{[m+3]} = N^{[m+3]}, \\
&\cdots \\
&\{N_0, F^{[2m-2]}\} + \sum_{j=0}^{m-3}\{N^{[m-j]}, F^{[m+j]}\} + R^{[2m-1]} = N^{[2m-1]}.
\end{aligned}
\tag{3.5}
$$

Recall that $2m_n - 1 = m_{n+1}$, see Sec. 2.1.2. From the formal solvability we know that each of these equations has a formal solution $F_n^{[m+j]}$. Of course, such a solution is not unique. We will make the solution unique by prescribing the condition

$$
\int_{\mathbb{T}^d} F_n^{[m+j]}(I, \theta) = 0.
$$

As we will see, this normalization will allow us to get the estimates needed for the proof of the convergence. The sum of the terms of orders $m_{n+1} + 1, \ldots, m_{n+1} + m_n - 2$ (i.e., $2m_n, \ldots, 3m_n - 3$) that appear in equation (4.1) is denoted by C_n. In the notation $m = m_n$, we have: $C_n = C_n^{[2m, 3m-3]}$. The terms of the uniform degree satisfy

$$
\begin{aligned}
&C_n^{[2m]} = \{N_{[3]}, F^{[2m-2]}\} + \{N_{[4]}, F^{[2m-3]}\} + \cdots + \{N_{[m]}, F^{[m+1]}\}, \\
&C_n^{[2m+1]} = \{N_{[4]}, F^{[2m-2]}\} + \{N_{[5]}, F^{[2m-3]}\} + \cdots + \{N_{[m]}, F^{[m+2]}\} \\
&\cdots \\
&C_n^{[3m-3]} = \{N_{[m]}, F^{[2m-2]}\}.
\end{aligned}
\tag{3.6}
$$

This can be written more compactly as

$$
C_n = \sum_{k=1}^{m-2}\{F^{[2m-1-k]}, \sum_{j=k+2}^{m} N^{[k+j]}\}.
\tag{3.7}
$$

This should be viewed as a definition of the remainder term C_n.

3.3 An important simplification.

In the case when the *normal form is an analytic function of $N_0(I)$* as in (2.2), we have an important simplification. Denote

$$
g_{2j}(I) := j b_j (N_0(I))^{j-1} \quad \text{and} \quad g_{2j+1}(I) \equiv 0.
\tag{3.8}
$$

Then for $j \in \mathbb{N}$ we have:

$$
\begin{aligned}
&\{N^{[2j]}, F\} = \{b_j(N_0)^j, F\} = j b_j (N_0)^{j-1}\{N_0, F\} = g_{2j}(I)\{N_0, F\} \\
&\{N^{[2j+1]}, F\} = g_{2j+1}(I)\{N_0, F\} \equiv 0.
\end{aligned}
\tag{3.9}
$$

We formulate this as a lemma:

Lemma 4. *If the normal form is an analytic function of $N_0(I)$ as in (2.2), then equation (3.5) is equivalent to*

$$\{N_0, F^{[m]}\} + R^{[m+1]} = N^{[m+1]},$$
$$\{N_0, F^{[m+1]}\} + g_3(I)\{N_0, F^{[m]}\} + R^{[m+2]} = N^{[m+2]},$$
$$\{N_0, F^{[m+2]}\} + g_4(I)\{N_0, F^{[m]}\} + g_3(I)\{N_0, F^{[m+1]}\} + R^{[m+3]} = N^{[m+3]},$$
$$\cdots$$
$$\{N_0, F^{[2m-2]}\} + \sum_{j=0}^{m-3} g_{m-j}(I)\{N_0, F^{[m+j]}\} + R^{[2m-1]} = N^{[2m-1]}.$$

$$(3.10)$$

and

$$C_n = \sum_{k=1}^{m-2} \left(\{F^{[2m-1-k]}, N_0\} \cdot \sum_{j=k+2}^{m} g_j \right). \tag{3.11}$$

3.4 Homological equations in majorants.

Here we study a simple recursive formula and estimate its terms. Later it will provide an important estimate of $|\{N_0, F^j\}|_{\rho_n, \rho_n}$. Here is the idea: suppose that in the Lemma above for some $\epsilon > 0$, for all $j = 0, \dots, m$ we have:

$$P_j := |R^{[m+j]}|_{\rho_n, \rho_n} + |N^{[m+j]}|_{\rho_n, \rho_n} \le \epsilon, \quad |g_j|_{\rho_n} \le 1/4^j.$$

Define S_j by the relations (3.12) below. Then, by Lemma 4, for all $j = 0, \dots, m$ we have

$$|\{N_0, F^j\}|_{\rho_n, \rho_n} \le S_j.$$

Lemma 5. *Given $\epsilon > 0$, suppose that for all $j = 1, \dots, m-1$ the numbers P_j satisfy*

$$0 < P_j \le \epsilon.$$

Let S_j be defined recursively by equations

$$S_1 = P_1,$$
$$S_2 = P_2 + \frac{1}{4}S_1,$$
$$S_3 = P_3 + \frac{1}{4}S_2 + \frac{1}{4^2}S_1$$
$$S_4 = P_4 + \frac{1}{4}S_3 + \frac{1}{4^2}S_2 + \frac{1}{4^3}S_1$$
$$\cdots$$
$$S_{m-1} = P_{m-1} + \sum_{j=1}^{m-1} \frac{1}{4^j}S_{m-1-j}.$$

$$(3.12)$$

Then for each j we have

$$S_j \le 2\epsilon, \quad j = 1, \dots, m-1.$$

Proof. By the formula for $S^{[j]}$ above,

$$S_j \leq P_j + \frac{1}{4}S_{j-1} + \frac{1}{4}(S_{j-1} - P_{j-1}) = P_j + 2\frac{1}{4}S_{j-1} \leq P_j + S_{j-1}/2.$$

This implies

$$S_j \leq \sum_{k=0}^{j-1} 2^{-k}P_{j-k} \leq \epsilon \sum_{k=0}^{j-1} 2^{-k} < 2\epsilon.$$

$\square$

4　Formal solution provides analytic with estimates.

In this section we study a homological equation (4.1) below with an analytic right-hand side $Q(I,\theta)$. Assuming that it has a formal solution, we will find an analytic one, and estimate it in terms of the right hand side. Similar procedures appear in [Ll].

Lemma 6. *Let $N_0(I) = I^{tr}\Omega I$ where Ω is a symmetric matrix with $\det \Omega \neq 0$, and let $Q(I,\theta)$ be analytic in an annulus $\mathbb{A}_{\rho,\sigma}$ for some ρ, $\sigma > 0$. Suppose that the following equation has a formal solution $\widetilde{F}(I,\theta)$:*

$$\{N_0, \widetilde{F}\} = Q. \tag{4.1}$$

Then equation (4.1) has an analytic solution $F(I,\theta)$, defined in $\mathbb{A}_{\rho,\sigma}$, and for any $0 < \delta < \rho$, $0 < \gamma < \sigma$ we have:

$$|F|_{\rho-\delta,\sigma-\gamma} \leq c(d,\Omega)\frac{1}{\delta\gamma^d}|Q|_{\rho,\sigma},$$

where $c(d,\Omega)$ is a constant only depending on d and Ω.

Moreover, if $Q(I,\theta)$ is a homogeneous polynomial in I with coefficients depending on θ, then so is $F(I,\theta)$.

Proof. Expanding F formally into a Fourier series: $F = \sum_{k \in \mathbb{Z}^d} \widehat{F}_k(I)e^{2\pi i\langle k,\theta\rangle}$, we get:

$$\{N_0, F\} = \sum_{j=1}^{d} F_{\theta_j}(N_0)_{I_j} = 2\pi i \sum_{k \in \mathbb{Z}^d} \langle k, 2\Omega I\rangle \widehat{F}_k(I)e^{2\pi i\langle k,\theta\rangle}.$$

Recall that Ω is symmetric, so $\langle k, \Omega I\rangle = \langle \Omega k, I\rangle$. Expressing $Q = \sum_{k \in \mathbb{Z}^d} \widehat{Q}_k(I)e^{2\pi i\langle k,\theta\rangle}$, we can rewrite equation (4.1) as a series of equations indexed by k:

$$\widehat{Q}_k(I) = 4\pi i\langle \Omega k, I\rangle\widehat{F}_k(I). \tag{4.2}$$

If $\langle k, \Omega I\rangle \neq 0$, we can express $\widehat{F}_k = \widehat{Q}_k(I)/(4\pi i\langle \Omega k, I\rangle)$.

Since we have assumed existence of a formal solution of the homological equation (4.1) (and hence, a solution of (4.2) for each k), we have:

$$\langle \Omega k, I\rangle = 0 \Rightarrow \widehat{Q}_k(I) = 0.$$

Hence, for $\langle \Omega k, I \rangle = 0$, the equation is satisfied for any value of $\widehat{F}_k(I)$. We define $\widehat{F}_k$ at these points by continuity. A way to do it is the following. Differentiate equation (4.2) in the direction of Ωk:

$$\langle \Omega k, \nabla \widehat{Q}_k(I) \rangle = 4\pi i \left(|\Omega k|^2 \widehat{F}_k(I) + \langle \Omega k, I \rangle \langle \Omega k, \nabla \widehat{F}_k(I) \rangle \right),$$

where for a vector $v \in \mathbb{R}^d$ we denote $|v|^2 = \sum_{j=1}^d v_j^2$. For $\langle \Omega k, I \rangle = 0$, define $\widehat{F}_k(I) = \langle \Omega k, \nabla \widehat{Q}_k(I) \rangle / (4\pi i |\Omega k|^2)$. Summing up, we have defined a continuous function $\widehat{F}_k(I)$ by

$$\widehat{F}_k(I) = \frac{1}{4\pi i} \begin{cases} \langle \Omega k, I \rangle^{-1} \widehat{Q}_k(I), & \langle \Omega k, I \rangle \neq 0, \\ \frac{1}{|\Omega k|^2} \langle \Omega k, \nabla \widehat{Q}_k(I) \rangle, & \langle \Omega k, I \rangle = 0. \end{cases}$$

Moreover, since $\widehat{F}_k(I)$ is analytic in $\mathbb{D}_\rho \setminus \{ \langle \Omega k, I \rangle = 0 \}$ and bounded in $\mathbb{D}_\rho$, it is analytic in $\mathbb{D}_\rho$. Notice that if in equation (4.2) $\widehat{Q}_k(I)$ is a homogeneous polynomial in I, then so is $\widehat{F}_k(I)$.

Now let us estimate the norm of the solution. Fix $0 < \delta < \rho/2$, $0 < \gamma < \sigma$. For each fixed $k \in \mathbb{Z}^d$, we will estimate the corresponding $\widehat{F}_k(I)$ in two steps: first "$\delta/2$-close" to the resonant plane $\langle \Omega k, I \rangle$, and then in the rest of $\mathbb{D}_{\rho-\delta}$.

For the first step, let $\Pi_\delta = \{ \langle \Omega k, I \rangle = 0 \} \cap \mathbb{D}_{\rho-\delta}$ be the part of the resonant plane falling into $\mathbb{D}_{\rho-\delta}$. Notice that the orthogonal complement to this plane is formed by the vectors $\alpha e^{2\pi i \phi} \Omega k$, $\alpha \geq 0$, $\phi \in [0, 1)$. Let

$$\Delta = \left\{ I = \alpha \frac{\Omega k}{|\Omega k|} e^{2\pi i \phi} \,\middle|\, \alpha < \delta/2, \, \phi \in [0, 1) \right\}$$

be the complex disk of radius $\delta/2$ centered at zero and orthogonal to Π_δ. Note that the restrictions of $\widehat{Q}_k(I)$ and $\widehat{F}_k(I)$ to this disc are analytic. Consider the $\delta/2$-neighbourhood O_δ of Π_δ: $O_\delta = \bigcup_{I_0 \in \Pi_\delta} (I_0 + \Delta)$. Then and $O_\delta \subset \mathbb{D}_{\rho-\delta}$.

For each fixed $I \in O_\delta$ there exists $I_0 \in \Pi_\delta$ such that $I \in I_0 + \Delta$. We can estimate $|\widehat{F}_k(I)|$ by the maximum modulus principle on the disk $I_0 + \Delta$. Namely, for I lying on the boundary of this disk we have: $|\langle \Omega k, I \rangle| = |\langle \Omega k, I_0 \rangle + \langle \Omega k, \delta \Omega k / (2|\Omega k|) \rangle| = |\Omega k| \delta/2$. Hence, for such I we have

$$|\widehat{F}_k(I)| \leq \frac{2|\widehat{Q}_k|_\rho}{4\pi \delta |\Omega k|} < \frac{|\widehat{Q}_k|_\rho}{\delta |\Omega k|}.$$

As the second step in this estimate, consider $I \in \mathbb{D}_{\rho-\delta} \setminus O_\delta$. Here $|\langle \Omega k, I \rangle| \geq |\Omega k| \delta/2$, so $|\widehat{F}_k(I)|$ satisfies the same estimate as above.

By Cauchy estimates, we have:

$$|\widehat{Q}_k|_\rho \leq |Q|_{\rho,\sigma} e^{-|k|\sigma}.$$

Since $\det \Omega \neq 0$, there exists a constant $c(\Omega)$ such that $|\Omega k| \geq |k|/c(\Omega)$ for all k. Then

$$|\widehat{F}_k|_{\rho-\delta} \leq \frac{1}{\delta |\Omega k|} |\widehat{Q}_k|_\rho \leq c(\Omega) \frac{e^{-\sigma|k|}}{\delta |k|} |Q|_{\rho,\sigma}.$$

Finally, for small δ and γ we have:

$$|F|_{\rho-\delta,\sigma-\gamma} \leq \sum_{k\in\mathbb{Z}^d\setminus\{0\}} e^{(\sigma-\gamma)|k|}|\widehat{F}_k|_{\rho-\delta} \leq \frac{c(\Omega)}{\delta} \sum_{k\in\mathbb{Z}^d\setminus\{0\}} \frac{e^{-\gamma|k|}}{|k|}|Q|_{\rho,\sigma}$$
$$\leq \frac{c(d,\Omega)}{\delta\gamma^d}|Q|_{\rho,\sigma},$$

where $c(d,\Omega)$ is a constant only depending on d and Ω. The estimates above are very wasteful, but they are enough for our purposes. $\qquad\square$

5   Proof of Proposition 1.

Here we summarize the preparatory work to complete the proof of Proposition 1. Let us return to the original problem. For a fixed n, let the necessary constants be as in Sec.2.1.2, $|\widetilde{R_n}|_{\rho_n} \leq \delta_n^\kappa$, and let $g_{2j}(I) = j\, b_j\, (N_0(I))^{j-1}$ as in (3.8).

5.1   Estimate of $|\{N_0, F_n\}|_{\rho_n,\rho_n}$ and $|C_n|_{\rho_n,\rho_n}$.

For $j = 1, \ldots, m_n - 1$ denote

$$P_j := |N^{[m_n+j]}|_{\rho_0} + |R^{[m_n+j]}|_{\rho_n}.$$

By the choice of ρ_0, see Sec. 2.1.2, for all $j = 1, \ldots, m_n - 1$ we have:

$$|g_j(I)|_{\rho_0} \leq 4^{-j}, \quad |N^{[m_n+j]}|_{\rho_0} \leq \delta_n^\kappa.$$

Since for $j = 1, \ldots, m_n - 1$ we have $|R^{[m_n+j]}|_{\rho_n} \leq |\widetilde{R_n}|_{\rho_n} \leq \delta_n^\kappa$, and so for these values of j we get

$$P_j \leq 2\delta_n^\kappa.$$

let S_j be defined by (3.12). By Lemma 5, for $j = 1, \ldots m - 1$ we have $S_j \leq 2\epsilon$. Equations (3.10) imply that for $j = 1, \ldots m - 1$ we have

$$|\{N_0, F_n^{[m+j-1]}\}|_{\rho_n,\rho_n} \leq S_j \leq 2\epsilon = 4\delta_n^\kappa. \tag{5.1}$$

By linearity,

$$|\{N_0, F_n\}|_{\rho_n,\rho_n} \leq \sum_{j=1}^{m_n-1} |\{N_0, F_n^{[m_n+j-1]}\}|_{\rho_n,\rho_n} \leq 4m_n\delta_n^\kappa \leq 4\delta_n^{\kappa-1}.$$

The latter estimate follows from the definition of m_n and δ_n, see Section 2.1.2.

Moreover, by (3.11),

$$|C_n|_{\rho_n} = \sum_{k=1}^{m-2}\left(S_{m-k}\sum_{j=k+2}^{m}G_j\right) \leq \sum_{k=1}^{m-2}\left(S_{m-k}\sum_{j=k+2}^{\infty}4^{-j}\right)$$
$$\leq \frac{1}{3}\sum_{k=1}^{m-2}4^{-(k+1)}S_{m-k} \leq \frac{1}{2}\epsilon = \delta_n^\kappa.$$

Hence,

$$|C_n|_{\rho_n} \leq \delta_n^\kappa. \tag{5.2}$$

5.2 Estimates for F_n.

Consider equation (5.1). Lemma 6 with $\rho = \sigma = \rho_n$, $\delta = \gamma = \delta_n$ and $|Q|_{\rho,\sigma} \leq 4\delta_n^\kappa$, implies:

$$|F_n^{[m+j-1]}|_{\rho_n - \delta_n, \rho_n - \delta_n} \leq 4c(d, \Omega)\delta_n^{\kappa - d - 1}.$$

Since $F_n = F_n^{[m_n, m_n + j - 1]}$ where $m_n \leq \delta_n^{-1}$, we get:

$$|F_n|_{\rho_n - \delta_n, \rho_n - \delta_n} \leq \sum_{j=1}^{m_n - 1} |F_n^{[m+j-1]}|_{\rho_n - \delta_n, \rho_n - \delta_n} \leq m_n \, 4c(d, \Omega)\delta_n^{\kappa - d - 1} \leq \delta_n^{\kappa - d - 3} \leq \delta_n^3. \quad (5.3)$$

The latter estimate follows from the definition of κ, see Section 2.1.2.

5.3 Estimates for Φ_n.

Here we prove that with F_n as above, estimates (2.10) and (2.11) hold true. Indeed, the coordinate change $\Phi_n = X_{F_n}^1$ is the time one map of the flow $X_{F_n}^t$ defined by the equations

$$\dot{I} = \partial_\theta F_n(I, \theta), \quad \dot{\theta} = -\partial_I F_n(I, \theta).$$

By (5.3) and Cauchy estimates we get

$$|\partial_I F_n|_{\rho_n - 2\delta_n, \rho_n - \delta_n} \leq \delta_n^2, \quad |\partial_\theta F_n|_{\rho_n - \delta_n, \rho_n - 2\delta_n} \leq \delta_n^2. \quad (5.4)$$

Then for any $t \leq 1$:

$$|X_{F_n}^t(I, \theta) - (I, \theta)|_{\rho_n - 3\delta_n, \rho_n - 3\delta_n} \leq t \, \delta_n^{-1}|F_n|_{\rho_n - 2\delta_n, \rho_n - 2\delta_n} \leq \delta_n^2.$$

$$X_{F_n}^t : \mathbb{A}_{\rho_n - 3\delta_n, \rho_n - 3\delta_n} \mapsto \mathbb{A}_{\rho_n - 2\delta_n, \rho_n - 2\delta_n} \quad (5.5)$$

In particular, since $\Phi_n = X_{F_n}^1$, we get the desired formulas (2.10) and (2.11).

5.4 Estimate of the new remainder $\widetilde{R_{n+1}}$.

Lemma 7. *For F_n constructed above, estimate (2.9) holds:*

$$|\widetilde{R_{n+1}}|_{\rho_n - 3\delta_n, \rho_n - 3\delta_n} < 4\delta_n^\kappa.$$

Proof. By Lemma 3,

$$\widetilde{R_{n+1}} = A_n + B_n + C_n,$$

where A_n, B_n and C_n are defined by (3.2) and (3.3).

Estimate of A_n: Using (5.5), we get:

$$|\widetilde{R_n}^{[>m_{n+1}]} \circ \Phi_n|_{\rho_n - 3\delta_n, \rho_n - 3\delta_n} \leq |\widetilde{R_n}|_{\rho_n - 2\rho_n, \rho_n - 2\delta_n} \leq \delta_n^\kappa.$$

Estimate of C_n: We showed in section 5.1 that

$$|C_n|_{\rho_n, \rho_n} \leq \delta_n^\kappa.$$

Estimate of B_n: By (5.4), $|\partial_I F_n|_{\rho_n - 2\delta_n, \rho_n - \delta_n} \leq \delta_n^2$ and $|\partial_\theta F_n|_{\rho_n - \delta_n, \rho_n - 2\delta_n} \leq \delta_n^2$. By (2.9)

$$|R_n|_{\rho_n, \rho_n} \leq |\widetilde{R_n}|_{\rho_n, \rho_n} \leq \delta_n^\kappa.$$

This implies, using Cauchy estimates, that

$$|\{R_n, F_n\}|_{\rho_n - 2\delta_n, \rho_n - 2\delta_n} \leq \delta_n^\kappa.$$

Notice that, by formulas (3.1) and (3.3), we have $\{N_n, F_n\} = R_n + N_n - N_{n-1} + C_n$. By (2.6),

$$|N_n - N_{n-1}|_{\rho_0, \rho_0} = \sum_{j=1}^{m_n} N^{[m_n + j]} \leq m_n \delta_n^{\kappa+1} \leq \delta_n^\kappa.$$

and therefore

$$|\{N_n, F_n\}|_{\rho_n, \rho_n} = |R_n|_{\rho_n, \rho_n} + |N_n - N_{n-1}|_{\rho_n, \rho_n} + |C_n|_{\rho_n, \rho_n} \leq 3\delta_n^\kappa.$$

Combining the above estimates, we get

$$|\{\{N_n, F_n\}, F_n\}|_{\rho_n - 2\delta_n, \rho_n - 2\delta_n} \leq \delta_n^\kappa,$$

Since, by (5.5), for any $t \leq 1$ we have $X_{F_n}^t : \mathbb{A}_{\rho_n - 3\delta_n, \rho_n - 3\delta_n} \mapsto \mathbb{A}_{\rho_n - 2\delta_n, \rho_n - 2\delta_n}$, we obtain

$$|\{\{N_n, F_n\} + R_n, F_n\} \circ X_{F_n}^t|_{\rho_n - 3\delta_n, \rho_n - 3\delta_n} \leq |\{\{N_n, F_n\} + R_n, F_n\}|_{\rho_n - 2\delta_n, \rho_n - 2\delta_n} \leq 2\delta_n^\kappa.$$

$\square$

Here we get the desired estimate for the remainder term. We have proved above that

$$|\widetilde{R_{n+1}}|_{\rho_n - 3\delta_n, \rho_n - 3\delta_n} < 4\delta_n^\kappa$$

Recall that $\widetilde{R_{n+1}} = \widetilde{R_{n+1}}^{[>m_{n+1}]}$. By Lemma 8 proved below, this implies the desired estimate

$$|\widetilde{R_{n+1}}|_{\rho_{n+1}, \rho_{n+1}} < \delta_{n+1}^\kappa$$

This finishes the proof of Proposition 1, and hence Theorem 1 (as explained in the introduction). $\square$

Lemma 8. *Suppose that the constants κ, b, δ_n, q_n, ρ_n are defined in Section 2.1.2, and an analytic function $G(I, \theta)$ satisfies $G = G^{[>m_{n+1}]}$, and*

$$|G|_{\rho_n - 3\delta_n, \rho_n - 3\delta_n} < 4\delta_n^\kappa.$$

Then

$$|G|_{\rho_{n+1}, \rho_{n+1}} < \delta_{n+1}^\kappa.$$

Proof. By the definition of κ in Section 2.1.2 we have: $q_n^{m_{n+1}+1} = q_n^{2^{n+1}+2} < q_n^{2^{n+1}} = 2b = 2^{-\kappa-2}$. Also recall that $\delta_{n+1} = 2^{-1}\delta_n$.

Since G starts with terms of degree $m_{n+1} = 2^{n+1} + 2$, we have:

$$|G|_{q_n(\rho_n - 3\delta_n), q_n(\rho_n - 3\delta_n)} < q_n^{2^{n+1}+2} 4\delta_n^\kappa \leq 2^{-\kappa-2} 4\delta_n^\kappa \leq \delta_{n+1}^\kappa.$$

$\square$

References

[Br71]   A. D. Bruno, *Analytic form of differential equations. I, II,* Tr. Mosk. Mat. Obs., 25, MSU, M., 1971, 119-262

[Br89]   A. D. Bruno, *Normalization of a Hamiltonian system near an invariant cycle or torus,* Russian Mathematical Surveys, 1989, 44:2, 53-89

[Dragt]   *Lie Methods for nonlinear Dynamics with applications to accelerator physics* https://www.physics.umd.edu/dsat/docs/Book19Nov2020.pdf.

[EFK13]   L.H. Eliasson, B. Fayad, R. Krikorian, *KAM-tori near an analytic elliptic fixed point,* Regular and Chaotic Dynamics, 18 no. 6, pp. 806-836 (2013)

[EFK15]   L.H. Eliasson, B. Fayad, R. Krikorian, *Around the stability of KAM tori,* Duke Math. Journ., 164 (2015), no 9, 1733-1775

[Ga]   G. Gallavotti, A criterion of integrability for perturbed harmonic oscillators. "Wick ordering" of the perturbations in classical mechanics and invariance of the frequency spectrum, Comm. Math. Phys. 87, 365-383.

[Gu66]   F. G. Gustavson, *On constructing formal integrals of a Hamiltonian system near an equilibrium point,* Astron. Jour., 71 (1966), pp. 670–686.

[Kri]   R. Krikorian, *On the divergence of Birkhoff Normal Forms,* preprint (2019); arXiv.org > math > arXiv:1906.01096

[Ll]   R. de la Llave, *On necessary and sufficient conditions for uniform integrability of families of Hamiltonian systems,* International Conference on Dynamical Systems (Montevideo, 1995), Pitman Res. Notes Math. Ser. vol. 362, pp. 76–109, Longman, Harlow, 1996

[LlMM]   R. de la Llave, J. Marco, R. Moriyón, *Canonical perturbation theory of Anosov systems and regularity results for the Livšic cohomology equation,* Annals of Mathematics,1986, pp. 537–611.

[MHO]   K. R. Meyer, and G. R. Hall, and D. Offin, *Introduction to Hamiltonian dynamical systems and the N-body problem,* Applied Mathematical Sciences, vol. 90, Springer-Verlag, New York, 2009.

[Mo60]   J. Moser, *On the integrability of area preserving Cremona mappings near an elliptic fixed point,* Bol. Soc. Mat. Mexicana (2), vol. 5 (1960), pp. 176–180

[Mu]   J. Murdock, *Normal forms and unfoldings for local dynamical systems,* Springer Monographs in Mathematics, Springer-Verlag, New York, 2003

[PM]   R. Pérez-Marco, *Convergence or generic divergence of the Birkhoff normal form,* Ann. of Math. (2), 157 (2003), no. 2, 557-574.

[Po92]   H. Poincaré, *Les méthodes nouvelles de la mécanique céleste. Tome I,* Librairie Scientifique et Technique Albert Blanchard, (1987)

[Rü59]   H. Rüssmann, *Über die Existenz einer Normalform inhaltstreuer elliptischer Transformationen,* Math. Ann., 137 (1959), pp.64–77

[Rü67]   H. Rüssmann, *Über die Normalform analytischer Hamiltonscher Differential-gleichungen in der Nähe einer Gleichgewichtslösung,* Math. Ann., 169 (1967), pp. 55–72

[Rü02]   H. Rüssmann, *Stability of elliptic fixed points of analytic area-preserving map-pings under the Bruno condition,* Ergodic Theory Dynam. Systems, 22 (2002) nr. 5, pp. 1551–1573

[Rü04]   H. Rüssmann, *Convergent transformations into a normal form in analytic Hamiltonian systems with two degrees of freedom on the zero energy surface near degenerate elliptic singularities,* Ergodic Theory Dynam. Systems, 24, (2004) nr. 5, pp. 1787–1832

[S]   M. Saprykina, *Domain of analyticity of normalizing transformations,* Nonlin-earity, 19 (2006), nr. 7, pp. 1581–1599

[Si54]   C.L. Siegel, *Über die Existenz einer Normalform analytischer Hamiltonscher Differentialgleichungen in der Nähe einer Gleichgewichtslösung,* Math. Ann., 128 (1954), pp. 144–170

[SM71]   C.L. Siegel and J. Moser, *Lectures on celestial mechanics,* Translation by Charles I. Kalme, Die Grundlehren der mathematischen Wissenschaften, Band 187 Springer-Verlag, New York, (1971)

[Ze73]   E. Zehnder, *Homoclinic points near elliptic fixed points* Comm. Pure Appl. Math., 26 (1973), pp.131–182

E-mail addresses:

R. de la Llave: `rafael.delallave@math.gatech.edu`

M. Saprykina: `masha@kth.se`