HAMILTONIAN NORMAL FORMS

MAURICIO GARAY AND DUCO VAN STRATEN

Abstract. We study a new type of normal form at a critical point of an analytic Hamiltonian. Under a Bruno condition on the frequency, we prove a convergence statement. Using this result, we deduce the existence of a positive measure set of invariant tori near the critical point.

To the memory of J.-C. Yoccoz.

Introduction

Investigations into normal forms of Hamiltonian systems can be traced back the earliest beginnings of celestial mechanics and perturbation theory in the works of Euler, Laplace, Delaunay, and others. The *Birkhoff normal form* provides a practical way to extend the classical theory of action-angle coordinates to critical points of Hamiltonians. As a general rule, the map which reduces the Hamiltonian to normal form is divergent, while under conditions of integrability it is convergent for analytic Hamiltonians ([4, 9, 24, 39, 29, 32, 34, 35, 36, 37]).
On the other hand KAM theory always provides, under non-degeneracy conditions, the existence of invariant tori. Since the appearance of Kolmogorov's original paper, the non-degeneracy conditions have been weakened and even in case the system is degenerate, invariant tori are known to exist ([1, 3, 5, 6, 17, 21, 27, 31, 33, 34]).

Here we present an iteration scheme that leads to a different type of normal form that we call the *Hamiltonian normal form*. It appears to be suited for the application of KAM theory near critical points of a Hamiltonian. We explain its relation to the classical Birkhoff normal form. From a formal perspective, the Hamiltonian normal form seems to be a rather trivial variant of the usual one, but it has some important technical advantages. The first advantage is that the Cantor set over which we work is *fixed* during the iteration, whereas in the usual proofs this set is constructed step by step and changes along the transformations of the iteration. A second advantage of the new iteration scheme is the control over what we call the *frequency space*. As a consequence, non-degeneracy conditions will be fulfilled automatically when passing to the limit. For the readers acquainted with Arnold's proof of the KAM theorem, this corresponds to the fact that the UV-cutoff can be chosen so that it involves only the terms that appear in the linear approximation.

The structure of the paper is as follows: In §1, after setting up some notations and a quick review of the classical Birkhoff normal form, we introduce the *frequency space* associated to a non-resonant critical point of a Hamiltonian. It is the space defined by the linear relations between the components of the gradient of the Birkhoff normal form. This space will play an important role and takes over the role of the usual non-degeneracy conditions of KAM theory.
In §2 we describe in some detail the formal aspects of an iteration that leads to our Hamiltonian normal form. The first important point is that the normalising maps in the iteration preserve, in a precise sense, the frequency space.
In §3 we state and prove, under a Bruno condition on the frequency, the convergence of the iteration procedure over a Cantor-like set in the complement of the *resonance fractal*. We introduce the relevant function spaces and rewrite the iteration in this context. The estimates we need are all simple applications of lemmas that are collected in an appendix.
In §4 we give an application to invariant tori near an elliptic fixed point. By our control of the frequency space, we can apply the arithmetic density theorem from [15] and obtain a measure result for the preimage under the frequency map. A conjecture formulated by M. Herman [18] is an easy corollary. The setup chosen here represents a significant improvement and simplification of the original arguments used in [11, 13, 14].

At the background of our investigations is a more general theory of normal forms based on a functional analytic theory of Banach space valued functors that we are currently developing. The reader interested in these lines of thought may take a look in the preliminary [8, 10, 13, 16]. From that perspective there appeared to be a certain incongruency between the usual Birkhoff normal form and the general theory of normal forms. The Hamiltonian normal form presented here seems to be the more natural one: the proof is from an abstract point of view in complete congruence with that of the ordinary Kolmogorov theorem. Nevertheless, for the convenience of the reader, we decided to keep the paper self-contained and provide complete proofs of all results used, with the exception of the arithmetical density result, which is taken from [15].

Acknowledgement. When we started to develop an abstract version of KAM theory, J.-C. Yoccoz was among the few enthusiastic dynamicists, eager to bridge the frontiers between algebra, topology and analysis.

After two months of numerous exchanges, he was forced to stop due to health problems. Yoccoz made several influential and motivating remarks, and for this reason we dedicate this research to his memory.

1. The Birkhoff normal form and the frequency space

We will be concerned with the structure of an analytic Hamiltonian system with d degrees of freedom near a critical point of the form

$$H = \sum_{i=1}^{d} \alpha_i p_i q_i + O(3).$$

We assume that the *frequency vector*:

$$\alpha := (\alpha_1, \alpha_2, \dots, \alpha_d) \in \mathbb{C}^d$$

is *non-resonant*, i.e., its components α_i are $\mathbb{Q}$-linearly independent. We can consider the Hamiltonian H as an element of the formal power series ring

$$P := \mathbb{C}[[q,p]] := \mathbb{C}[[q_1, \dots, q_d, p_1, \dots, p_d]].$$

The Poisson bracket of $f, g \in P$, defined by

$$\{f, g\} = \sum_{i=1}^{d} \partial_{q_i} f \partial_{p_i} g - \partial_{p_i} f \partial_{q_i} g,$$

makes P into a Poisson-algebra. An element $h \in P$ is a power series that can be written as

$$h := \sum_{a,b} C_{a,b} p^a q^b, \quad C_{a,b} \in \mathbb{C},$$

where we use the usual multi-index notation, so that

$$p^a q^b := p_1^{a_1} p_2^{a_2} \dots p_d^{a_d} q_1^{b_1} q_2^{b_2} \dots q_d^{b_d},$$

and so on. We will assign weight 1 to each of the variables, so that the monomial $p^a q^b$ has weight $|a| + |b|$. We write $h = O(k)$ if h only contains monomials of degree $\geq k$, and say that h *has order* k. If h is analytic, it is represented by a convergent series, and our usage of the O corresponds to its usual meaning.

1.1. ***Birkhoff normal form.*** A derivation $v \in Der(P)$ that preserves the Poisson-bracket:

$$v(\{f, g\}) = \{v(f), g\} + \{f, v(g)\}$$

is called a *Poisson-derivation* and we denote by $\Theta(P)$ the vector space of all Poisson-derivations or Poisson vector fields. The map

$$P \longrightarrow \Theta(P), \quad h \mapsto \{-, h\}$$

associates to h the corresponding Poisson-derivation, usually called the *Hamiltonian vector field* of h. If $h = O(k)$ and $f = O(l)$, then clearly

$\{f,h\} = O(k+l-2)$, so the vector field $v := \{-,h\}$ is said to be *of order* $k-2$, although the coefficients of the vector field v are $O(k-1)$. In particular, if $h = O(3)$, then one can *exponentiate* v and obtain a Poisson automorphism of the ring P:

$$e^v = Id + \{-,h\} + \frac{1}{2!}\{\{-,h\},h\} + \ldots \in Aut(P).$$

If we let

$$h_0 := \sum_{i=1}^{d} \alpha_i p_i q_i,$$

then

$$\{h_0, p^a q^b\} = (\alpha, a-b) p^a q^b,$$

where $(-,-)$ denotes the standard euclidean scalar product. So if α is non-resonant, then each monomial $p^a q^b$ with $a \neq b$ appearing in $H = h_0 + O(3)$ can be removed by an application of the derivation

$$v = j(p^a q^b) := \{-, \frac{1}{(\alpha, a-b)} p^a q^b\}.$$

Hence, we can construct a sequence of automorphisms

$$\varphi_0 := e^{-v_0}, \quad \varphi_1 := e^{-v_1}, \quad \varphi_2 := e^{-v_2}, \ldots, \in Aut(P)$$

that remove successively all monomials $p^a q^b$, $a \neq b$ of weight $k+3$ from H_k, defined recursively by

$$H_1 := \varphi_0(H),\ H_2 = \varphi_1(H_1),\ \ldots,$$

so that the automorphism

$$\Phi_k := \varphi_{k-1} \ldots \varphi_1 \varphi_0$$

maps H to H_k. The infinite composition

$$\Phi := \ldots \varphi_k \varphi_{k-1} \ldots \varphi_1 \varphi_0 \in Aut(P)$$

is a formal symplectic coordinate transformation that removes all monomials $p^a q^b$, $a \neq b$ from our Hamiltonian H, so that

$$\Phi(H) = B_H,$$

where B_H is a series of the form

$$B_H := \sum_{a \in \mathbb{N}^d} C_a p^a q^a.$$

The series B_H is called the *Birkhoff normal form* of H. There exist several variants of this algorithm, which differ in details and notation. For example, it is possible to remove certain terms at the same time. These may lead to different normalising transformations Φ, but it is known that these different choices lead to the same series B_H. As in the iteration process one has to divide by the quantity $(\alpha, a-b)$ to remove the $p^a q^b$-term, small denominators appear, which lead to convergence problems for the formal series involved.

C. L. Siegel has shown that for a generic critical point of a real analytic Hamiltonian H with $d \geq 2$ degrees of freedom, the series of the transformation Φ is in fact divergent [35]. It is expected that in general the series B_H itself is also divergent, but no published proof is known to us (see [29] for more details).

1.2. **The Moser Extension.** As the monomials $p_i q_i$ $(i = 1, 2, \ldots, d)$ Poisson commute with the Birkhoff normal form B_H, Birkhoff normalisation implies that any non-resonant Hamiltonian H is *formally completely integrable.* To express this fact more clearly, it is useful to enlarge the ring P and consider

$$Q := \mathbb{C}[[\tau, q, p]] = \mathbb{C}[[\tau_1, \ldots, \tau_d, q_1, \ldots, q_d, p_1, \ldots, p_d]]$$

with the extra τ-variables, introduced by Moser. With the same definition of the Poisson-bracket as before, Q becomes a Poisson algebra with Poisson centre $Q_0 := \mathbb{C}[[\tau]]$. We will assign weight $= 2$ to the variables τ_i, so that the d elements

$$f_i := p_i q_i - \tau_i \in Q$$

are homogeneous of degree two. These elements Poisson commute, $\{f_i, f_j\} = 0$, and we obtain a Poisson commuting sub-algebra

$$\mathbb{C}[[\tau, f]] = \mathbb{C}[[\tau, f_1, f_2, \ldots, f_d]] = \mathbb{C}[[\tau, p_1 q_1, \ldots, p_d q_d]]$$

containing Q_0. The $f_1, f_2, \ldots, f_d$ also generate an ideal

$$I = (f_1, f_2, \ldots, f_d) \subset Q = \mathbb{C}[[\tau, q, p]]$$

and clearly, the canonical map

$$\mathbb{C}[[\tau, p, q]] \longrightarrow \mathbb{C}[[p, q]], \quad p_i \mapsto p_i,\ q_i \mapsto q_i,\ \tau_i \mapsto p_i q_i.$$

induces an isomorphism of the factor ring Q/I with our original ring P:

$$Q/I \xrightarrow{\sim} P.$$

Although f_i maps to zero under this map, the derivation $\{-, f_i\}$ induces the non-zero derivation $\{-, p_i q_i\}$ on P, so the map $Q \longrightarrow P$ is *not* a Poisson-morphism. The ideal $I^2 \subset Q$ is the square of the ideal I, i.e. generated by the elements $f_i f_j$, $1 \leq i, j \leq d$, and plays a very distinguished role in dynamics. The reason is that if $T \in I^2$, then $\{h, T\} \subset I$. As a consequence, H and $H + T$ determine the *same Hamiltonian vector field* on $Q/I = P$.

Extending the multi-index notation in an obvious way, we can write

$$p^a q^a = (\tau + f)^a = \tau^a + \sum_{i=1}^{d} \partial_{\tau_i} \tau^a f_i + I^2.$$

The term τ^a is in the centre of Q, whereas the above remark implies that $p^a q^a$ and $\sum_{i=1}^{d} \partial_{\tau_i} \tau^a f_i$ define the same derivation on the ring $P = Q/I$.

We can consider the Birhoff normal form series $B(pq) = B_H(pq)$ as an element of Q. When we write $pq = \tau + f$, then we find:

$$B(\tau + f) = B(\tau) + \sum_{i=1}^{d} b_i(\tau) f_i \bmod I^2.$$

The first term $B(\tau)$ belongs to the Poisson centre Q_0 and is dynamically trivial, but gets mapped to the non-trivial element $B_H \in P$. The second term $\sum_{i=1}^{d} b_i(\tau) f_i$ carries the dynamical information in Q, but is mapped by the canonical map $Q \longrightarrow P$ to zero.

The formal power series $b_1, \dots, b_d \in \mathbb{C}[[\tau]]$ are obtained as partial derivatives of B, considered as a series in the τ_i-variables:

$$b = (b_1, \dots, b_d) = \nabla B(\tau).$$

One has $b(0) = \alpha$, and the higher order terms describe how the frequencies change with τ and for this reason we call it the (formal) *frequency map*. If the system happens to be integrable, then the series are convergent and the vector $b(\tau) = (b_1(\tau), b_2(\tau), \dots, b_d(\tau))$ is the frequency of motion on the corresponding manifold defined by $f_i(\tau, q, p) = 0$, $i = 1, 2, \dots, d$.

For a multi-index $a \in \mathbb{N}^d$, we consider the vector

$$\nabla^a b(0) := \frac{\partial^a b}{\partial \tau^a}(0) \in \mathbb{C}^d.$$

Definition 1.1. *The* frequency space *of H is the vector space $F(H) \subset \mathbb{C}^d$ generated by the vectors $\nabla^a b(0)$, $|a| \geq 1$.*

This vector space controls an important aspect of the non-degeneracy conditions in KAM theory. The space $F(H) = \{0\}$ if there are no terms of degree > 2 in the Birkhoff normal form. This is the case considered by Rüßmann [32], who proved, under arithmetic conditions on the frequency vector α, that in this situation the transformation Φ actually is *convergent*, so H is analytically completely integrable. The classical non-degeneracy condition from KAM-theory lead to the opposite case $F(H) = \mathbb{C}^d$. Our main interest lies in the intermediate cases. In §4 we will define in the real domain a C^∞-map

$$\beta : \mathcal{W} \longrightarrow \alpha + \mathcal{F}(H) \subset \mathbb{R}^d,$$

defined in a neighbourhood $\mathcal{W} \subset \mathbb{R}^d$, $\mathcal{F}(H) := F(H) \cap \mathbb{R}^d$ and whose Taylor series coincides with the above formal power series b.

2. The Hamiltonian normal form

We will now describe a variant of the Birkhoff normal form algorithm that allows for a better control of the invariant tori. For this we have to introduce d further additional variables

$$\omega_1, \omega_2, \ldots, \omega_d,$$

to which we assign weight zero. The iteration will take place in the formal power series ring with $4d$ variables

$$R := \mathbb{C}[[\omega, \tau, q, p]] = \mathbb{C}[[\omega_1, \omega_2, \ldots, \omega_d, \tau_1, \ldots, \tau_d, q_1, \ldots, q_d, p_1, \ldots, p_d]].$$

Again we retain the standard Poisson bracket, so now the Poisson centre is the subring $R_0 := \mathbb{C}[[\omega, \tau]]$. The following sub-algebra is of importance for our discussion:

Definition 2.1. *We let*

$$M := \mathbb{C}[[\omega, \tau, f]] = \mathbb{C}[[\omega, \tau, pq]] \subset R,$$

and call it the Moser-algebra *of R.*

We remark that M is a Poisson commutative sub-algebra of R and that all monomials of M have even degree. We will also make use of the projection

$$\pi : R \longrightarrow M$$

that maps all monomials not in M to zero. Clearly, $\pi \circ i = Id_M$, where $i : M \longrightarrow R$ is the inclusion.

2.1. ***The homological equation.*** As before, we denote the vector space of Poisson derivations of R by $\Theta(R)$, which has the structure of a module over the Poisson centre R_0. These derivations decompose into Hamiltonian and non-exact parts:

$$\Theta(R) = \mathrm{Ham}\,(R) \oplus \mathrm{Der}\,(R_0),$$

that is, an element of $\Theta(R)$ is of the form

$$v = \{-, h\} + w$$

with

$$w = \sum_{i=1}^{d} A_i \frac{\partial}{\partial \omega_i} + B_i \frac{\partial}{\partial \tau_i}, \quad A_i, B_i \in R_0.$$

Definition 2.2. *Let $F, m \in R$ and $v \in \Theta(R)$. If*

$$v(F) = m,$$

we say that v solves the homological equation *for m on F.*

We will now solve the homological equation for certain special elements. The following function will play an important role in the paper.

Definition 2.3. *The* formal unfolding *of $h_0 = \sum_{i=1}^{d} \alpha_i p_i q_i$ is the element*

$$A_0 := \sum_{i=1}^{d} (\alpha_i + \omega_i) p_i q_i \in R.$$

So A_0 is obtained from h_0 by detuning the frequencies in the most general way.

The infinitesimal action

$$\Theta(R) \longrightarrow R, \quad v \mapsto v(A_0)$$

on A_0 takes a simple form in the monomial basis:

$$\begin{aligned} \{A_0, p^a q^b\} &= (\alpha+\omega, a-b) p^a q^b, \\ \partial_{\omega_k} A_0 &= p_k q_k, \\ \partial_{\tau_k} A_0 &= 0. \end{aligned}$$

Definition 2.4. *We define a* $\mathbb{C}[[\omega,\tau]]$*-linear map*

$$L : R \longrightarrow \Theta(R) = \mathrm{Ham}\,(R) \oplus \mathrm{Der}\,(R_0), m \mapsto Lm$$

by setting for $a \neq b$*:*

$$Lp^a q^b := \{-, \frac{1}{(\alpha+\omega, a-b)} p^a q^b\}.$$

For $a = b$*, or more generally for a series*

$$m = g(p_1 q_1, p_2 q_2 \ldots, p_d q_d) = g(pq)$$

we set

$$Lm := \sum_{i=1}^{d} \frac{\partial g(\tau)}{\partial \tau_i} \partial_{\omega_i}.$$

We note that the first case of the definition applies to the monomials in the kernel of the projection map π, whereas the second part of the definition applies to the elements of the Moser sub-algebra M.

Definition 2.5. *For* $A = A_0 + T,\ T \in I^2$ *we define a linear map*

$$j_A : R \longrightarrow \Theta(R)$$

in terms of L *by the formula*

$$j_A : m \mapsto Lm - L(Lm(T)) = L(m - Lm(T))$$

Proposition 2.6. *For any* $A = A_0 + T \in A_0 + I^2$*, there exists* $t \in R_0 + I^2$ *such that*

$$j_A(m)(A) = m + t.$$

Proof. First, for $A = A_0$ we have $j_{A_0} = L$. For $m = p^a q^b$ with $a \neq b$ we have

$$j_{A_0}(m)(A_0) = \{A_0, \frac{1}{(\alpha+\omega, a-b)} p^a q^b\} = p^a q^b = m$$

and for $m = g(pq)$ we have, with $g_i = \partial_{\tau_i} g$,

$$\begin{aligned} j_{A_0}(m)(A_0) &= \sum_{i=1}^{d} g_i(\tau)\frac{\partial A_0}{\partial \omega_i} = \sum_{i=1}^{n} g_i(\tau)p_i q_i \\ &= \sum_{i=1}^{d} g_i(\tau) f_i \bmod R_0 = g(pq) \bmod R_0 + I^2, \end{aligned}$$

where we used the Taylor expansion

$$g(pq) = g(\tau + f) = g(\tau) + \sum_{i=1}^{d} g_i(\tau) f_i \bmod I^2.$$

This shows the correctness for $T = 0$. For the general case $A = A_0 + T$, we get

$$\begin{aligned} j_A(m)(A_0 + T) &= Lm(A_0) + Lm(T) - L(Lm(T))A_0 - L(Lm(T))(T) \\ &= m + Lm(T) - Lm(T) - L(Lm(T))(T) \bmod R_0 + I^2 \\ &= m + L(Lm(T))(T) \bmod R_0 + I^2. \end{aligned}$$

Because $T \in I^2$, it follows that $Lm(T) \in I$. Furthermore, for any $g \in I$, we have $Lg(T) \in I^2$. This can be seen by writing g as $\mathbb{C}[[\omega,\tau]]$-linear combination of terms of the form $p^a q^b f_i$. If $a \neq b$, $\{T, p^a q^b f_i\} \in I^2$, whereas for $a = b$, we obtain a combination of terms $\partial_{\omega_i} T$, which is in I^2, as the generators $f_i = p_i q_i - \tau_i$ are independent of ω_i. □

2.2. ***Hamiltonian normal form iteration.*** Our Hamiltonian normal form is obtained from the following basic iteration. Starting from a Hamiltonian

$$H = \sum_{i=0}^{d} \alpha_i p_i q_i + O(3),$$

we first form

$$F_0 := H + \sum_{i=1}^{d} \omega_i p_i q_i = A_0 + O(3).$$

Solving the homological equation on A_0 for the degree 3 part of F_0 determines a Poisson derivation v_0. The application of e^{-v_0} to F_0 produces F_1, where this term is removed and we put $A_1 = A_0$. The solution of the homological equation for the degree 4 and 5 part of F_1 on A_1 defines v_1. Then the application of e^{-v_1} to F_1 produces F_2, where now the terms of degree 4 and 5 are removed, modulo terms in $(R_0 + I^2) \cap M$. These terms we add to $A_1 = A_0$ and obtain A_2. Then we solve the homological equation for the terms of degree $6, 7, 8, 9$ of

F_2, but now on A_2, etcetera. Thus we obtain a by iteration a sequence of triples

$$(F_n, A_n, v_n), \quad n = 0, 1, 2, \dots$$

that we call the *Hamiltonian normal form iteration* and which we will describe now in some detail. It is convenient to write

$$[h]_i^j$$

for the sum of terms of h of weight $\geq i$ and $< j$, so that $[h]_i^{i+1}$ represents the part of h of pure weight i. When $j = +\infty$ we omit the letter j, when $i = 0$ we omit the letter i. In a similar way we can truncate a vector field by truncating its coefficients, but taking the shift of grading by 1 into account.

We begin with the *initialisation step*

$$\begin{aligned} F_0 &= H + \sum_{i=1}^{d} \omega_i p_i q_i = A_0 + O(3) \\ A_0 &= \sum_{i=1}^{d} (\alpha_i + \omega_i) p_i q_i \\ v_0 &= [j_{A_0}\left([F_0]_3^4\right)]^2. \end{aligned}$$

The next terms are determined by the *iteration step*:

From F_n, A_n we first compute

$$v_n = [j_{A_n}([F_n]_{2^n+2}^{2^{n+1}+2})]^{2^{n+1}}.$$

and then obtain

$$\begin{aligned} F_{n+1} &= e^{-v_n} F_n, \\ A_{n+1} &= A_n + [F_n - v_n(F_n)]_{2^n+2}^{2^{n+1}}. \end{aligned}$$

Then compute

$$v_{n+1} := [j_{A_{n+1}}([F_{n+1}]_{2^{n+1}+2}^{2^{n+2}+2})]^{2^{n+2}}$$

etcetera. The extra truncation of v_n is not needed for the arguments here, but it will play a role in 2.6.

It is useful to define the *increments*

$$S_{n+1} := [F_n - v_n(F_n)]_{2^n+2}^{2^{n+1}+2},$$

so that:

$$A_{n+1} = A_n + S_{n+1}.$$

There are a few simple but important points to notice:

Proposition 2.7.
(i) The derivation v_n has order 2^n, i.e. $v_n = [v_n]_{2^n} = [v_n]_{2^n}^{2^{n+1}}$.
(ii) $F_n = A_n + O(2^n + 2)$.
(iii) $S_n \in (R_0 + I^2) \cap M$.

Proof. (i) From the recursive definition we see that v_n is obtained by solving the homological equation with the terms of degrees $2^n + 2$ up to $2^{n+1} + 2$ from F_n. Taking Poisson-bracket with a term of degree $2^n + 2$ shifts degrees by 2^n, and similarly for the non-exact part of v_n. So indeed v_n has order 2^n.
(ii) This follows from an easy induction on n. By definition, the statement holds for $n = 0$. Let us assume that

$$F_n = A_n + O(2^n + 2)$$

From the definition of F_{n+1} we have

$$F_{n+1} = e^{-v_n} F_n = F_n - v_n(F_n) + \frac{1}{2} v_n^2(F_n) - \dots$$

and as v_n has order 2^n, it follows that

$$v_n^2(F_n) = O(2 + 2^n + 2^n) = O(2^{n+1} + 2).$$

So we have

$$F_{n+1} = A_n + [F_n - v_n(F_n)]_{2^n+2}^{2^{n+1}+2} + O(2^{n+1} + 2) = A_{n+1} + O(2^{n+1} + 2).$$

(iii) We use induction on n and assume that $S_n \in (R_0 + I^2) \cap M$. From (ii) we have

$$F_n = A_n + O(2^n + 2).$$

The derivation $v_n = [j_{A_n}([F_n]_{2^n+2}^{2^{n+1}+2})]_{2^n}^{2^{n+1}}$ is constructed to solve the homological equation up to terms of high order:

$$v_n(A_n) = [F_n]_{2^n+2}^{2^{n+1}+2} + t + O(2^{n+1} + 2), \quad t \in (R_0 + I^2) \cap M$$

As we have

$$v_n(F_n) = v_n(A_n + O(2^n + 2)) = v_n(A_n) + O(2^{n+1} + 2),$$

we see that the increment

$$[F_n - v_n(F_n)]_{2^n+2}^{2^{n+1}+2} \in (R_0 + I^2) \cap M,$$

hence also $S_{n+1} \in (R_0 + I^2) \cap M$. □

2.3. **Quadratic and non-quadratic nature of the iteration.** The Hamiltonian normal form iteration produces a sequence (F_n, A_n, v_n): the series

$$F_0 = H + \sum_{i=1}^{d} \omega_i p_i q_i = A_0 + O(3)$$

is transformed by

$$\Phi_n := e^{-v_{n-1}} \dots e^{-v_0}$$

to a series of the form

$$F_n = A_n + O(2^n + 2).$$

If we let n go to ∞, we obtain a formal Poisson automorphism

$$\Phi_\infty := \dots e^{-v_n} \dots e^{-v_0} \in Aut(R),$$

and obtain

$$F_\infty := \Phi_\infty(F_0) = A_\infty, \quad A_\infty \in A_0 + (R_0 + I^2) \cap M$$

The automorphism Φ_∞ transforms the perturbation $F_0 = A_0 + O(3)$ back to the normal form A_0, plus terms that have no effect on the dynamics.

For convenience of the reader we include the following diagram that indicates the degrees of the quantities that appear in the iteration.

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
F_0	$\bullet$	$\times$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$
F_1	$\bullet$	$\circ$	$\times$	$\times$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$
F_2	$\bullet$	$\circ$	$\bullet$	$\circ$	$\times$	$\times$	$\times$	$\times$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$	$\blacksquare$
F_3	$\bullet$	$\circ$	$\bullet$	$\circ$	$\bullet$	$\circ$	$\bullet$	$\circ$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\blacksquare$
F_4	$\bullet$	$\circ$	$\bullet$	$\circ$	$\bullet$	$\circ$	$\bullet$	$\circ$	$\bullet$	$\circ$	$\bullet$	$\circ$	$\bullet$	$\circ$	$\bullet$	$\circ$	$\times$

The bullets $\bullet$ and circles $\circ$ represent terms of A_n. They belong to the $R_0 + I^2$ part of the Moser-algebra : the $\bullet$ terms are constant in columns, the circles $\circ$ are zero, as the Moser-algebra only has terms of even degree.
So $\bullet$ and $\circ$ represents the *normal form range*, consisting of terms of F_n of degree

$$2 \leq degree < 2^n + 2$$

The crosses $\times$ represent the terms of F_n that determine the derivations v_n. These make up what we call the *active range* of degrees:

$$2^n + 2 \leq degree < 2^{n+1} + 2.$$

The black squares $\blacksquare$ represent the terms of F_n that of degree higher than $2^{n+1} + 2$ that do not directly influence the next iteration step, but of course must be carried along.

We now rewrite the iteration in a form where this trichotomy in degrees is manifest. Consider the decomposition

$$F_n := A_n + M_n + U_n = \bullet + \times + \blacksquare,$$

where

$$A_n := [F_n]^{2^n+2}, \quad M_n := [F_n]_{2^n+2}^{2^{n+1}+2}, \quad U_n := [F_n]_{2^{n+1}+2},$$

are the lower, middle and upper part of F_n. The middle term decomposes, by the construction of v_n, as sum of two terms

$$(1)\quad M_n = S_{n+1} + [v_n(A_n)]_{2^n+2}^{2^{n+1}+2},$$

where

$$S_{n+1} = [F_n - v_n(F_n)]_{2^n+2}^{2^{n+1}+2} \in (R_0 + I^2) \cap M,$$

where we also used

$$[v_n(A_n)]_{2^n+2}^{2^{n+1}+2} = [v_n(F_n)]_{2^n+2}^{2^{n+1}+2}$$

The quantity $v_n(A_n)$ has only terms of middle and high degree:

$$(2)\quad v_n(A_n) = [v_n(A_n)]_{2^n+2}^{2^{n+1}+2} + R_n,$$

where

$$R_n := [v_n(A_n)]_{2^{n+1}+2}.$$

represents a remainder part. We then have

$$\begin{aligned}
F_{n+1} &= e^{-v_n} F_n \\
&= e^{-v_n}(A_n + M_n + U_n) \\
&\stackrel{1}{=} e^{-v_n}(A_n + S_{n+1} + [v_n(A_n)]_{2^n+2}^{2^{n+1}+2} + U_n) \\
&\stackrel{2}{=} e^{-v_n}(A_n + S_{n+1} + v_n(A_n)) + e^{-v_n}(U_n - R_n) \\
&= A_n + S_{n+1} + (e^{-v_n} - \mathrm{Id}\,)(A_n + S_{n+1}) + e^{-v_n}(v_n(A_n)) + e^{-v_n}(U_n - R_n) \\
&= A_n + S_{n+1} + (e^{-v_n}(\mathrm{Id}\, + v_n) - \mathrm{Id}\,)A_n + (e^{-v_n} - \mathrm{Id}\,)(S_{n+1}) + e^{-v_n}(U_n - R_n)
\end{aligned}$$

where we indicated the equations we used to establish the equalities. The third and fourth equalities are just obtained by rearrangement of the different terms.

Now, when we set

$$B_n := M_n + U_n = [F_n]_{2^n+2} = \times + \blacksquare,$$

we can formulate the *KAM-form of the iteration in the formal setting:*

$$\boxed{\begin{aligned}
S_{n+1} &= [B_n - v_n(A_n)]_{2^n+2}^{2^{n+1}+2}, \\
A_{n+1} &= A_n + S_{n+1}, \\
B_{n+1} &= \psi(v_n)S_{n+1} + \phi(v_n)A_n + e^{-v_n}([B_n - v_n(A_n)]_{2^{n+1}+2})
\end{aligned}}$$

with

$$\psi(z) = e^{-z} - 1, \quad \phi(z) = e^{-z}(1+z) - 1$$

and where we used the abbreviation:

$$v_n = j_n B_n = [j_{A_n}(B_n)]_{2^n}^{2^{n+1}}.$$

Note that the increments S_{n+1} and the vector fields v_n are auxiliary quantities, which depend linearly on B_n. We see that the terms involving ϕ and ψ are quadratic in B_n, whereas the third term is not. But that term only depends on the higher order terms of the series.

In the classical Kolmogorov scheme, the convergence is ensured by the quadraticity of the iteration, while in the Arnold-Nash-Moser case there does appear such a remainder term from the classical UV cutoff technique. However, one may hope that these terms will remain small, as they involve terms of sufficiently high order.
As we will see, unlike the cases considered by these authors, in our situation it is important to be very precise about the degrees in the truncation, as otherwise we would loose control over the fields v_n and the convergence properties of the iteration could be spoiled.

2.4. **Some further remarks on the iteration.** Let us denote by

$$R^{an} := \mathbb{C}\{\omega, \tau, p, q\}$$

the Poisson sub-algebra of R consisting of *convergent power series*. Its Poisson centre is $R_0^{an} := \mathbb{C}\{\omega, \tau\}$. If $v \in \Theta(R^{an})$ is an analytic Poisson-derivation of order ≥ 1, then the exponential series e^v converges as an element of $Aut(R^{an})$. As a result, if we start with an analytic Hamiltonian, the terms of iteration (F_k, S_k, v_k) are are all analytic, but of course the limit objects a priori are given by formal series.
The variable ω plays a very special role. From the definition of the operation L, it follows that the iteration can be formulated in a much smaller ring. We define the ring SD_α of *small denominators* at α as the subring of the field $\mathbb{C}(\omega)$ of rational functions, defined by localisation of $\mathbb{C}[\omega]$ with respect with the multiplicative subset S generated by all linear polynomials $(\alpha + \omega, J)$, $J \in \mathbb{Z}^n \setminus \{0\}$:

$$SD_\alpha := \mathbb{C}[\omega]_S := \mathbb{C}[\omega, \frac{1}{(\alpha + \omega, J)}, J \in \mathbb{Z}^n \setminus \{0\}] \subset \mathbb{C}(\omega).$$

Then the iteration makes sense in the Poisson-algebra

$$R^{an} \cap SD_\alpha[[\tau, p, q]],$$

but we will not make use of this fact in this paper.

In section 3 we will formulate a version of this iteration in appropriate function spaces and study its convergence properties.

2.5. **Relation to the Birkhoff normal form.** When we start from an analytic Hamiltonian $H(p,q)$, we first formed

$$F_0(\omega,p,q)=H(p,q)+\sum_{i=1}^{d}\omega_i p_i q_i,$$

and the Hamiltonian normal form iteration produces a sequence of Poisson derivations:

$$v_n,\quad n=0,1,2,\ldots$$

These exponentiate to Poisson-automorphisms $\varphi_n := e^{-v_n}$ and the compositions

$$\Phi_n=\varphi_{n-1}\varphi_{n-1}\ldots\varphi_1\varphi_0,\quad \varphi_n:=e^{-v_n},$$

preserve the ring $R^{an}\subset R$ of convergent series. We have

$$\Phi_n(F_0)=F_n\in\mathbb{C}\{\omega,\tau,p,q\},$$

and the infinite composition is a Poisson automorphism of the formal power series ring R:

$$\Phi_\infty=\ldots\varphi_{n-1}\varphi_{n-1}\ldots\varphi_1\varphi_0\in Aut(R),$$

and correspondingly

$$\Phi_\infty(F_0)=F_\infty\in\mathbb{C}[[\omega,\tau,p,q]].$$

Note that

$$F_\infty=A_\infty=A_0+T_\infty,\quad T_\infty\in R_0+I^2.$$

From the construction of the vector fields v_n we see that

$$\Phi_n(\tau_i)=\tau_i,\ \ \Phi_\infty(\tau_i)=\tau_i.$$

As any Poisson-automorphism, Φ_n and Φ_∞ preserve the Poisson centre, so we have

$$\Phi_n(\omega_i)\in\mathbb{C}\{\omega,\tau\},\ \ \Phi_\infty(\omega_i)\in\mathbb{C}[[\omega,\tau]].$$

As these element will play an important role in the sequel, will give the them a fixed name:

Definition 2.8. *We set*

$$R_{n,i}:=\Phi_n(\omega_i)\in\mathbb{C}\{\omega,\tau\},$$
$$R_{\infty,i}:=\Phi_\infty(\omega_i)\in\mathbb{C}[[\omega,\tau]].$$

To see what happens to the original Hamiltonian $H(p,q)$ during the iteration, we remark that

$$F_0(\omega=0,\tau,p,q)=H(p,q).$$

But during the iteration, the condition $\omega_i=0$ is transformed into the condition $R_{n,i}(\omega,\tau)=0$:

$$\Phi_n(F_0)=F_0(R_n(\omega,\tau),\tau,P_n(\omega,\tau,p,q),Q_n(\omega,\tau,p,q))=F_n(\omega,\tau,p,q).$$

So if we want to follow $H(p,q)$ during these transformations, we will need to solve for ω the equations

$$R_{n,i}(\omega,\tau)=0,\ \ i=1,2,\dots,d,$$

and

$$R_{\infty,i}(\omega,\tau)=0,\ \ i=1,2,\dots,d.$$

From the fact that vector fields v_n have order 2^n, we readily see that

$$R_{0,i}=\omega_i,\quad R_{n,i}=R_{n-1,i}+O(2^n),$$

so the equations $R_{n,i}(\omega,\tau)=0$ can be solved for the ω_i and we obtain convergent power series

$$\omega_{n,i}(\tau)\in\mathbb{C}\{\tau\},$$

such that

$$R_{n,i}(\omega_n(\tau),\tau)=0,$$

and similarly for $n=\infty$, we find formal power series

$$\omega_{\infty,i}(\tau)\in\mathbb{C}[[\tau]]$$

solving

$$R_{\infty,i}(\omega_\infty(\tau),\tau)=0.$$

So we then have

$$\Phi_n(H(p,q))=F_n(\omega_n(\tau),\tau,p,q)=:H_n(\tau,p,q).$$

and is thus of the form

$$H_n=h_n(\tau)+\sum_{i=1}^d(\alpha_i+\omega_{n,i}(\tau))f_i+O(2^n+2)\ \mathrm{mod}\, I^2\cap M$$

with $f_i=p_iq_i-\tau_i$.

Comparing with the Taylor expansion of 1.2:

$$B(qp)=B(\tau+f)=B(\tau)+\sum_{i=1}^d b_i(\tau)f_i\ \mathrm{mod}\, I^2,$$

we deduce that the coefficients $b_i(\tau)$ of the frequency map are related to the $\omega_{n,i}(\tau)$ by the congruence

$$\alpha_i+\omega_{n,i}(\tau)=b_i(\tau)+O(2^n+2),$$

and that the constant term gives back the Birkhoff normal form of H:

$$h_n(\tau)=B(\tau)+O(2^n+2).$$

One may be tempted to go one step further and use the projection $Q\longrightarrow P$, where we eliminate the τ-variables using the relations $f_i=0$, $i=1,2,\dots,d$, in order to create an iteration taking place in the ring P. This certainly can be done, but the resulting formulas are much more complicated and less revealing. The great advantage of working with the Hamiltonian normal form iteration is that we *avoid* solving these implicit relations $R_{n,i}(\omega,\tau)=0$ and $f_i=0$. This leads to a great simplification of the setup.

2.6. **Behaviour of the frequency space.** We will now see that the precise truncations that we made for the derivations v_n allow us to control some non degeneracy properties from the formal iteration.
We can find open subsets $U_n\subset\mathbb{C}^{2d}$ in the $2d$-dimensional ω,τ space, on which the power series $R_{n,i}(\omega,\tau),\ \omega_{k,i}(\tau),\ i=1,2,\dots,n$ converge. They define d-dimensional complex analytic manifolds

$$X_n := \{(\tau,\omega)\in U_n \mid R_{n,1}(\tau,\omega)=\cdots=R_{n,d}(\tau,\omega)=0\},$$

which are the graphs of the map

$$\omega_n=(\omega_{n,1}(\tau),\dots,\omega_{n,d}(\tau)).$$

Proposition 2.9. *The manifolds X_n are contained in $F(H)\times\mathbb{C}^d$, where $F(H)$ denotes the frequency space of H.*

Proof. We prove the statement by induction on n. Clearly, this is true for X_0, which reduces to the d-dimensional plane

$$\omega_1=\cdots=\omega_d=0.$$

As X_{n+1} is the image of X_n under the automorphism by $\varphi_n=e^{-v_n}$, it is sufficient show that the restriction of the vector field v_n to X_n is contained in the frequency space $F(H)$.

To see this, decompose the field v_n into Hamiltonian and non-Hamiltonian part

$$v_n=\sum_{i=1}^d v_{i,n}(\omega,\tau)\partial_{\omega_i}+w_n,\quad w_n\in\mathrm{Ham}\,(Q).$$

Only the non-Hamiltonian part enters the computation of

$$\varphi_n(\omega_i)=\omega_i-v_{n,i}(\omega,\tau)+O(2^k+2).$$

As the $R_{n,i}(\omega,0)=\omega_i+O(2)$, we may apply Weierstrass division and divide the coefficients $v_{n,i}$ successively by the $R_{n,j}$'s and get an expression of the form:

$$v_{n,i}=a_{n,i}(\tau)+\sum_{j=1}^d R_{n,j}(\omega,\tau)b_{n,j}(\omega,\tau).$$

By definition, the coefficients $R_{n,j}(\tau,\omega)$ vanish on X_n. The Hamiltonian $H_{n+1}=(F_{n+1})_{|X_{n+1}}$ is in Birkhoff normal form up to order $2^{n+1}+1$ and moreover a direct computation shows that:

$$\begin{aligned}H_{n+1}&=(A_n+v_n(A_n))_{|X_{n+1}}+O(2^{n+1}+2)\\&=(A_0)_{|X_{n+1}}+\sum_{i=1}^d a_{n,i}(\tau)p_iq_i+O(2^{n+1}+2)\bmod R_0+I^2\end{aligned}$$

Therefore by identification with the Birkhoff normal form, we get that

$$a_{n,i}(\tau) = [b_{n,i}(\tau)]_{2^n}^{2^{n+1}}.$$

This shows that the vector field v_n is tangent to the frequency space along X_n. Thus its time one flow e^{-v_n} maps the analytic manifold $X_n \subset F(H) \times \mathbb{C}^d$ to an analytic submanifold $X_{n+1} \subset F(H) \times \mathbb{C}^d$. This proves the proposition. □

This simple proposition will turn out to be a fundamental property of the formal Hamiltonian normal form iteration.

3. Convergence of the Hamiltonian normal form

We now proceed to the convergence properties of the iteration scheme of the Hamiltonian normal form. Our strategy will be to transfer the Hamiltonian normal form from power series rings to function spaces and "replace" the ring R by a sequence of Banach spaces E_n, where the E_n are certain spaces of functions on a shrinking sequence of sets

$$W_0 \supset W_1 \supset W_2 \supset \ldots \supset W_\infty,$$

so that there will be natural restriction mappings

$$E_0 \hookrightarrow E_1 \hookrightarrow E_2 \hookrightarrow \ldots \hookrightarrow E_\infty$$

of norm ≤ 1. The iteration step will construct from $A_n, B_n \in E_n$ elements $A_{n+1}, B_{n+1} \in E_{n+1}$, using certain operators that emulate the operators that appear in the iteration scheme of §2.

It is important to have a certain flexibility in the possible choices of the spaces E_n. The maps occurring in the iteration will always satisfy precise estimates. The resulting technical difficulties are just of notational nature and largely irrelevant, as we shall see.

3.1. **The resonance fractal.** The ω-variables are crucial: we can control the estimates only if we restrict the ω-domain over a sequence of decreasing sets. In this way we obtain convergence over the remaining Cantor-like set.

Let us denote by

$$\Xi_v := \{\beta \in \mathbb{C}^d \mid (\beta, v) = 0\} \subset \mathbb{C}^d$$

the hyperplane orthogonal to vector $v \in \mathbb{C}^d, v \neq 0$, with respect to the standard scalar product

$$(\beta, v) := \sum_{i=1}^{d} \beta_i v_i.$$

A vector $\beta \in \mathbb{C}^d$ is *non-resonant* precisely if for all $J \in \mathbb{Z}^d \setminus \{0\}$

$$(\beta, J) \neq 0.$$

Definition 3.1. *The* complex resonance fractal *is the set*

$$\Xi := \bigcup_{J\in\mathbb{Z}^d\setminus\{0\}} \Xi_J,$$

The real resonance fractal *is the set*

$$\mathcal{X} := \bigcup_{J\in\mathbb{Z}^d\setminus\{0\}} \mathcal{X}_J,$$

where $\mathcal{X}_J = \Xi_v \cap \mathbb{R}^d$.

Proposition 3.2. i) *The set* $\Xi \subset \mathbb{C}^d$ *is dense for* $d > 2$.
ii) *The set* $\mathcal{X} \subset \mathbb{R}^d$ *is dense for* $d > 1$.

Proof. We only prove the first assertion. The second one is similar, but easier. As $\mathbb{Q}$ is dense in $\mathbb{R}$, it follows that Ξ is dense in the set

$$S := \{\beta \in \mathbb{C}^d : \exists \;\; v \in \mathbb{R}^d \setminus \{0\}, (\beta, v) = 0\},$$

consisting of all points the belong to the complexification of a real hyperplane. Let us show that for $d > 2$, we have $S = \mathbb{C}^d$. Let $z \in \mathbb{C}^d = \mathbb{R}^d + i\mathbb{R}^d$ and write

$$z = x + iy, \;\; x, y \in \mathbb{R}^d.$$

As $d > 2$, there exists a real hyperplane $V \subset \mathbb{R}^d$ containing the vectors $x, y \in \mathbb{R}^d$, so

$$z = x + iy \in V \otimes \mathbb{C}.$$

This proves the statement. □

3.2. **Arithmetical classes.** In general, if β is non-resonant, the scalar product (β, J), $J \in \mathbb{Z}^d \setminus \{0\}$ is non-zero, but can become arbitrary small. To quantify this, we consider the following.

Definition 3.3. *The* arithmetic sequence $\sigma(\beta) = (\sigma(\beta)_k)$ *of a vector* $\beta \in \mathbb{C}^d$ *is defined as*

$$\sigma(\beta)_k := \min\{|(\beta, J)| : J \in \mathbb{Z}^d \setminus \{0\}, \|J\| \leq 2^k\}.$$

The precise nature of the falling sequence $\sigma(\beta)$ encodes important arithmetical properties of the vector β.

Definition 3.4. *Let* $a = (a_k)$ *be any real positive decreasing sequence. The* complex arithmetic class *associated to* a *is the set*

$$\mathcal{C}(a)_\infty := \cap_{m=0}^{\infty} \mathcal{C}(a)_m, \text{ where } \mathcal{C}(a)_m := \{\beta \in \mathbb{C}^n : \forall k \leq m \;\; \sigma(\beta)_k \geq a_k\}.$$

Similarly, the real arithmetic class *associated to* a *is the set*

$$\mathcal{R}(a)_\infty := \cap_{m=0}^{\infty} \mathcal{R}(a)_m \;\text{ where }\; \mathcal{R}(a)_m := \{\beta \in \mathbb{R}^d : \forall k \leq m \;\; \sigma(\beta)_k \geq a_k\}.$$

Note that

$$\mathcal{R}(a)_m = \mathcal{C}(a)_m \cap \mathbb{R}^d,\ \forall m \in \mathbb{N} \cup \{\infty\}.$$

If a vector β belongs to $\mathcal{C}(a)_\infty$, then $|(\beta, J)| \geq a_k$ for all lattice vectors $0 \neq \|J\| \leq 2^k$, and thus small denominators are controlled by the sequence a.

As we are dealing with a descending chain of closed sets $\mathcal{C}(a)_m$ and $\mathcal{R}(a)_m$, it follows that $\mathcal{C}(a)_\infty$ and $\mathcal{R}(a)_\infty$ are closed subsets. From the above propositions we can conclude immediately:

Corollary 3.5. i) *The set $\mathcal{C}(a)_\infty$ has an empty interior for $d > 2$.*
ii) *The set $\mathcal{R}(a)_\infty$ has an empty interior for $d > 1$.*

So the case of two degrees of freedom is special. In this case, the complex arithmetic classes $\mathcal{C}(a)_\infty \subset \mathbb{C}^2$ have a non-empty interior. More precisely, if the frequency ratio is non-real,

$$\beta_2/\beta_1 \notin \mathbb{R},$$

then it is an interior point of $\mathbb{C}^2 \setminus \Xi$. This ensures a convergent Birkhoff normalisation in two degrees of freedom. This peculiar fact was already observed by Moser back in 1958 [25] (see also [28]).

Although we tautologically have

$$\beta \in \mathcal{R}(\sigma(\beta))_\infty,$$

it might be an isolated point of that set.

Note however, that $a \geq a'$ then $\mathcal{C}(a)_\infty \subset \mathcal{C}(a')_\infty$ and similarly $\mathcal{R}(a)_\infty \subset \mathcal{R}(a')$. The following elementary but fundamental result shows that although arithmetic classes may have an empty interior, after replacing a by a slightly smaller sequence νa, they are *big* in the sense of measure theory.

Proposition 3.6 ([15])**.** *Let $a = (a_n)$ be a positive decreasing sequence and $\beta \in \mathcal{R}(a)_\infty$. Let $\nu = (\nu_n)$ be another positive decreasing sequence with $\nu_i \leq 1$ and $\sum_{i=1}^\infty \nu_i < \infty$. Then the density of $\mathcal{R}(\nu a)_\infty$ at β is equal to 1:*

$$\lim_{\varepsilon \longrightarrow 0} \frac{\mu(B(\beta, \varepsilon) \cap \mathcal{R}(\nu a)_\infty)}{\mu(B(\beta, \varepsilon))} = 1$$

Here νa is the sequence with terms $\nu_n a_n$ and μ denotes the Lebesgue measure, $B(\beta, \varepsilon)$ the ball with radius ε, centred at β.

Definition 3.7. *For a given vector $\beta \in \mathbb{R}^d$, we say that a falling sequence $a = (a_n)$ is β-*dense*, if the set $\mathcal{R}(a)_\infty$ has density 1 at the point β:*

$$\lim_{\varepsilon \longrightarrow 0} \frac{\mu(B(\beta, \varepsilon) \cap \mathcal{R}(a)_\infty)}{\mu(B(\beta, \varepsilon))} = 1$$

So the above proposition says we can always find $a < \sigma(\beta)$ slightly smaller, such that a is β-dense. Similar statements holds in the complex case be we will not use it, as our range of applications is in the real domain.

3.3. **The sets Z_n and W_n.** We will have to study the set $\mathcal{C}(a)_\infty$ in the neighbourhood of a fixed frequency vector α. For this we consider a decreasing sequence $s = (s_n)$ and converging to a positive limit s_∞. We denote by $B(r) \subset \mathbb{C}^d$ the ball of radius r centred at the origin.

Definition 3.8. *For fixed decreasing sequences a and s we define the closed set*

$$Z_n := Z_n(\alpha, a, s) := \{\omega \in B(s_n) : \forall k \leq n, \sigma(\alpha + \omega)_k \geq a_k(s_0 - s_n)\}$$

Note that $Z_0 = B(s_0)$. As the sequence (s_n) is decreasing we have

$$a_k(s_0 - s_{n+1}) \geq a_k(s_0 - s_n),$$

so the sets (Z_n) form a descending chain and can be considered as a local variant of the chain $(\mathcal{C}(a)_n)$. Moreover if $s_0 < 1$ then $a_k > a_k(s_0 - s_n)$ for all n's and therefore

$$B(s_n) \cap \mathcal{C}(a)_n \subset Z_n$$

In the iteration, we will have to control the shrinking of the sets Z_n. To do so, given two open sets $V \subset U \subset \mathbb{C}^d$, we denote by $\delta(U, V)$ the supremum of the real numbers ρ for which

$$V + \rho B \subset U.$$

where B denotes the unit ball.

Lemma 3.9. *The sequence $(\delta(Z_n, Z_{n+1}))_{n \in \mathbb{N}}$ satisfies the estimate*

$$\delta(Z_n, Z_{n+1}) \geq 2^{-n} a_n(s_n - s_{n+1}).$$

Proof. The proof is straightforward. Assume that $\omega \in Z_{n+1}$ and take $x \in \mathbb{C}^d$ satisfying

$$\|x\| \leq 2^{-n} a_n(s_n - s_{n+1}).$$

For $k \leq n$ and $\|J\| < 2^k$, we have:

$$\begin{aligned} |(\alpha + \omega + x, J)| &\geq |(\alpha + \omega, J)| - |(x, J)| \\ &\geq a_k(s_0 - s_{n+1}) - \|x\| \, \|J\| \\ &\geq a_k(s_0 - s_{n+1}) - a_n(s_n - s_{n+1}) \\ &\geq a_k(s_0 - s_{n+1}) - a_k(s_n - s_{n+1}) = a_k(s_0 - s_n). \end{aligned}$$

This shows that $\omega + x \in Z_n$ and thus proves the lemma. □

The sets W_n we will be working with, are defined as follows.

Definition 3.10. *For fixed decreasing sequences a and s we set*
$$W_n := Z_n \times D^d_{s_n} \times D^{2d}_{s_n} \subset \mathbb{C}^d \times \mathbb{C}^d \times \mathbb{C}^{2d}$$
where $D_t \subset \mathbb{C}$ denotes the closed disc of radius t. Furthermore, we put
$$V_n := Z_n \times D^d_{s_n} \subset \mathbb{C}^d \times \mathbb{C}^d$$

The coordinates on $\mathbb{C}^d \times \mathbb{C}^d \times \mathbb{C}^{2d}$ are
$$\omega_1, \dots, \omega_d,\ \tau_1, \dots, \tau_d,\ q_1, \dots, q_d,\ p_1, \dots, p_d,$$
and there are projection maps
$$W_n \longrightarrow V_n, \quad (\omega, \tau, q, p) \mapsto (\omega, \tau).$$

3.4. **The Banach spaces.** The Banach spaces that we will use, consist of functions that are holomorphic on the interior of the closed sets W_n defined above. They come in different flavours: we may require that they extend continuously to the boundary, or more generally be Whitney C^k, or be square integrable.

For a closed subset $X \subset \mathbb{R}^d$, we denote by
$$C^k(X, \mathbb{R})$$
the vector space of C^k*-Whitney differentiable functions*, $k \in \mathbb{N}$, for which the norm
$$\|f\| = \max_{|I| \leq k} \sup_{x \in X} |\partial^I f(x)|$$
is finite and therefore defines a Banach space structure on it[1].

There is a natural definition of a holomorphic function at a point x of a closed subset $X \subset \mathbb{C}^k$. Let us say that a vector v at $x \in X$ is *interior* if:
$$\exists \delta, |t| < \delta \implies x + tv \in X.$$
Then f *is holomorphic at $x \in X$ if for any interior vector v with the property the limit*
$$\lim_{t \mapsto 0} \frac{f(x + tv) - f(x)}{t}$$
exists for complex values of t.

If x is an interior point this is equivalent to the standard definition. In the general case it allows us to control analyticity of the function on the

[1]Recall that a function $f : X \longrightarrow \mathbb{R}$ defined on a closed subset X of Euclidean space is called *Whitney differentiable* at $x \in X$, if there exists functions $D^I f(x)$ called the Whitney derivatives of f at x such that:
$$f(y) = \sum_{|I| \leq m} \frac{D^I f(x)}{|I|!} (y - x)^I + o(\|y - x\|^m)$$
The sole difference that the Whitney definition bears is the uniformity of the limit in the x, y variables.

boundary of a set and also to avoid defining our spaces as topological tensor products.

If f is defined in a closed polydisc $D_1 \times D_2$, to be holomorphic means that

(1) f is holomorphic in the interior of the polydisc.
(2) $f(x_1, -)$ and $f(-, x_2)$ define holomorphic functions for any x_1 and any x_2.

It follows from Morera's theorem that if f is continuous then the first condition implies the second one. On the other extreme if f is defined in $X = \{0\} \times D_2$, then to be holomorphic means that $f(0, -)$ is holomorphic. In this case, the set X has an empty interior. In our iteration the limit set over which the functions are defined, namely W_∞ has indeed an empty interior while for finite n, the sets W_n are the closures of their interior.

Definition 3.11. *Given a closed subset $X \subset \mathbb{C}^d$, we denote by $\mathcal{O}^k(X)$ the Banach space of (bounded) C^k-Whitney function which are holomorphic in X:*

$$\mathcal{O}^k(X) = C^k(X, \mathbb{C}) \cap \mathcal{O}(X)$$

For $k = 0$, we use the notation $\mathcal{O}^c$ instead of $\mathcal{O}^k$.
We define similarly $\mathcal{O}^h(X)$ as the Hilbert space of square integrable functions on X, holomorphic on the interior.

For our iteration scheme, we will use the following Banach spaces:

Definition 3.12.

$$E_n^b := \mathcal{O}^b(W_n),\ b = h, c \text{ or } k,\ n \in \mathbb{N} \cup \{\infty\}.$$

Note that E_∞^b is the subspace of $\mathcal{O}^b(W_\infty)$ consisting of functions for which $f(\omega, -)$ is holomorphic for any $\omega \in Z_\infty$:

$$E_\infty^k = \{f \in C^k(W_\infty) : \forall \omega \in Z_n,\ f(\omega, -) \in \mathcal{O}^k(D_{s_n}^{3d})\}$$

That these subspaces are closed follows from Morera's theorem.

Also the spaces of functions

$$F_n^b := \mathcal{O}^b(V_n)$$

do play a role. Note that $\mathcal{O}^b(V_n) \subset \mathcal{O}^b(W_n)$, so F_n^k can be identified with the closed subspaces of E_n^k, consisting of those functions that *do not depend on q, p*.

The restriction mappings induce injective mappings

$$(E_n^b, |\cdot|_n) \longrightarrow (E_m^b, |\cdot|_m),\ m > n,$$

and this holds also for $m = \infty$, since the image consists of Whitney C^∞ functions having analytic Taylor expansion. As the set W_n is connected, this expansion defines uniquely a holomorphic function in this neighbourhood.

Usually there is no ambiguity to drop the n in $|x|_n$; for $x \in E_n$ the notation $|x|$ just means $|x|_n$. Furthermore, if (x_n) is a sequence with $x_n \in E_n$, then the notation $|x|$ can be used for the *sequence* $|x_n|$ and be called the *norm sequence* of x, etc.

We say that a sequence (x_n), $x_n \in E_n$ is *convergent*, if the image sequence in E_∞ converges and that it converges *quadratically*, if there is an estimate of the form

$$|x - x_n|_\infty \leq Cq^{2^n},$$

for some $q \in]0,1[$ and $C > 0$.

3.5. **Bruno sequences.** We will make use of Bruno sequences in two different ways. First, we will have to impose a Bruno-condition on the frequency vector α and second, we will use a Bruno sequence ρ to determine the sequence s of polyradii. This leads to a simple way to control the small denominator estimates for the operators that will be defined in the next section.

Definition 3.13 ([2]). *A strictly monotone positive sequence a is called a* Bruno sequence *if the infinite product*

$$\prod_{k=0}^{\infty} a_k^{1/2^k}$$

converges to a strictly positive number or equivalently if

$$\sum_{k \geq 0} \left| \frac{\log a_k}{2^k} \right| < +\infty.$$

Since their introduction in [2], these sequences have played a key role in KAM-theory. We denote respectively by $\mathcal{B}^+$ and $\mathcal{B}^-$ the set of increasing and decreasing Bruno sequences. The set of Bruno sequences some obvious multiplicative properties:

i) Taking the multiplicative inverse interchanges $\mathcal{B}^+$ and $\mathcal{B}^-$,
ii) The product of two elements in $\mathcal{B}^\pm$ is again in $\mathcal{B}^\pm$.
iii) An element of $\mathcal{B}^\pm$ raised to a positive power remains in $\mathcal{B}^\pm$.

Note also that any geometrical sequence $a_n = q^n$, $q \neq 1$ is a Bruno sequence, belonging to $\mathcal{B}^-$ if $q < 1$, to $\mathcal{B}^+$ if $q > 1$. The sequence $a_n = e^{\pm\alpha^n}$ belongs to $\mathcal{B}^\pm$ if and only if $1 < \alpha < 2$.

Definition 3.14. *A vector $\alpha \in \mathbb{C}^d$ is said to satisfy* Bruno's arithmetical condition *if $\sigma(\alpha) \in \mathcal{B}^-$.*

If $(a_n) \in \mathcal{B}^+, a_0 > 1$, then from the convergence of the sum $\sum \frac{\log a_n}{2^n}$ we see that for any $\varepsilon > 0$, there is an N such that for $n \geq N$ we have

$$\log a_n \leq 2^n \varepsilon, \quad a_n \leq (e^\varepsilon)^{2^n},$$

so the sequence can not increase faster than a quadratic iteration

$$u_{n+1} = u_n^2, \ u_0 > 1.$$

Similarly, for $(a_n) \in \mathcal{B}^-, a_0 < 1$ we find an estimate

$$a_n \geq (e^{-\varepsilon})^{2^n} \ \text{ for } \ n \geq N,$$

so that the sequence can not decrease too quickly.

Sequences of type $\mathcal{P}(a)$. We need to make a special choice of our sequence s that determines the size of the sets W_n in terms of an auxiliary Bruno sequence $\rho = (\rho_n) \in \mathcal{B}^-$.

We will always assume that $\rho_0 < 1$ and that ρ converges to 0. For a given $s_0 > 0$, we define s_n recursively by

$$s_{n+1} = (\rho_n)^{1/2^n} s_n.$$

Because of the Bruno property of ρ, the sequence $s = (s_n)$ converges to a positive limit s_∞.

The sequence ρ will be taken small enough to counteract certain small denominators of the form

$$a_n^l (s_n - s_{n+1})^m = a_n^l (1 - \rho^{1/2^n})^m s_n^m$$

that appear norm estimates.Therefore we define

Definition 3.15. *Given a sequence $a = (a_n) \in \mathcal{B}^-$, we denote by $\mathcal{P} = \mathcal{P}(a)$ (resp. $\mathcal{P}^+$) the set of sequences depending on parameters $\rho_n > 0$, $n \in \mathbb{N}$, whose terms are majorated by $C\rho_n^k a_n^{-l}(s_n - s_{n+1})^{-m}$ for some $C, k, l, m \geq 0$ (resp. $k > 0$):*

$$u_n \leq \frac{C\rho_n^k}{a_n^l (s_n - s_{n+1})^m}$$

The following lemma shows that an appropriate choice of ρ absorbs the small denominators.

Proposition 3.16. *For any falling Bruno sequence $b \in \mathcal{B}^-$ and any $u \in \mathcal{P}^+(a)$, there exists values of the sequence $\rho = \rho(a, b)$ and a constant $K \in \mathbb{R}_+$ so that $u < Kb$ and $\rho \in \mathcal{B}^-$.*

Proof. We let

$$u_n < C\rho_n^k a_n^{-l} (s_n - s_{n+1})^{-m},$$

and show that the falling Bruno sequence

$$\rho_n = 2^{(-n-1)m/k} M^{-1/k} a_n^{l/k} b_n^{1/k} \text{ with } M = \max(C, 2^k a_0^l b_0)$$

satisfies the condition.

For this choice, we have:

$$u_n < 2^{(-n-1)m} b_n (s_n - s_{n+1})^{-m}$$

By our choice of the sequence $s = (s_n)$ we have:

$$s_n - s_{n+1} = (1 - \rho_n^{1/2^n}) s_n \geq (1 - \rho_n^{1/2^n}) s_\infty.$$

The constant M is defined so that the falling sequence ρ is bounded by $1/2$ and hence:

$$(1 - \rho_n^{1/2^n}) \geq (1 - 2^{-1/2^n}) > 2^{(-n-1)}.$$

Thus we have the estimates

$$u_n < 2^{(-n-1)m} b_n (s_n - s_{n+1})^{-m} < b_n s_\infty^{-m} = K b_n.$$

This proves the proposition. □

So in the iteration process, small denominators $a_n^l (s_n - s_{n+1})^m$ can be *absorbed* by the presence of such a sequence ρ in the numerator, and in fact be pushed below any pre-given Bruno sequence.

3.6. **The operators.** The basic step in the iteration of §2 that determines A_{n+1}, B_{n+1} form A_n, B_n depended on the truncation operators

$$[-]_{2^n+1}^{2^{n+1}+1}, \quad [-]_{2^{n+1}+2}$$

and the map j_n, which came from a map $j_A : R \longrightarrow \Theta(R)$, which was defined as a composition of several other maps. If we want to lift these operator in the Banach space context, we often need to shrink the domain of definition slightly. For this reason it is convenient to define Banach spaces $E_{n+\varepsilon}$ between E_n and E_{n+1}, and set up things in such a way that the iteration step brings us from E_n to E_{n+1}, using one or more intermediate spaces $E_{n+\varepsilon}$, where $0 < \varepsilon < 1$. We implement this idea in the following simple way: we interpolate the sequence $s = (s_n)$ by setting:

$$s_{n+\varepsilon} = (\rho_n)^{\varepsilon/2^n} s_n, \ \varepsilon \in]0, 1[,$$

and then set

$$W_{n+\varepsilon} := Z_{n+\varepsilon} \times D^d_{s_{n+\varepsilon}} \times D^{2d}_{s_{n+\varepsilon}},$$
$$E^b_{n+\varepsilon} := \mathcal{O}^b(W_{n+\varepsilon}),$$

where we put

$$Z_{n+\varepsilon} = \{\omega \in \mathbb{C}^d \mid \forall k \leq n, \sigma(\alpha + \omega)_k \geq a_k (s_0 - s_{n+\varepsilon}\}.$$

One then has:

$$\delta(W_n, W_{n+\varepsilon}) \geq \frac{a_n}{2^n}(s_n - s_{n+\varepsilon}) \geq \frac{a_n}{2^n}(1 - \rho_n^{\varepsilon/2^n}) s_\infty.$$

The definition of $\mathcal{P}$ and $\mathcal{P}^+$ extend for a fixed ε as well as the absorption principle.

Restriction maps. First, there are the restriction operators $E^b_n \longrightarrow E^b_m$, $n \geq m$. These have norm ≤ 1 and thus are completely harmless. We will not mention or even give names to these maps. A different matter is that we also have to use 'flavour changing' restriction maps, like the canonical map

$$r_{n,m} : E^h_n \longrightarrow E^c_m, (n < m),$$

which arise by remarking that a square integrable holomorphic function on W_n is continuous on the smaller set W_m. From the local equivalence lemma (Lemma 5.2), it follows that the norm sequence of $r_{n,n+\varepsilon}$ belongs to $\mathcal{P}$. There is also a canonical map $i_n : E^c_n \longrightarrow E^h_n$, whose norm is bounded by the volume of W_n and as these volumes form a falling sequence, this norm sequence is also in $\mathcal{P}$.

Truncation maps. For the power series ring we used the truncation operations

$$[-]^{2^{n+1}+2}_{2^n+2}, \quad [-]_{2^{n+1}+2},$$

that pick out some range of monomials from a series. Inside the Hilbert spaces E^h_n the monomials form an orthogonal set on polydiscs, so these truncations are just special orthogonal projectors. In this way we obtain corresponding maps

$$\tau^h_{n,m} : E^h_n \longrightarrow E^h_m, \quad (n < m),$$

and

$$\sigma^h_{n,m} : E^h_n \longrightarrow E^h_m, \quad (n < m).$$

Corresponding truncation operators on E^c-spaces can then be defined as compositions

$$\tau^c_{n,m} : E^c_n \xrightarrow{i_n} E^h_n \xrightarrow{\tau^h_{n,m'}} E^h_{m'} \xrightarrow{r_{m',m}} E^c_m,$$

where $n < m' < m$. We can make the specific but arbitrary choice $m' = (m+n)/2$.

The norm of these maps can be estimated directly from the Arnold-Moser lemma (Lemma 5.6) and the conclusion is that the norm sequences for $\tau^b_{n,n+\varepsilon}$ and $\sigma^b_{n,n+\varepsilon}$ ($b \in \{h, c\}$) belong to $\mathcal{P}^+$ for $b = h$. Due to local equivalence (Lemma 5.2), it also belongs to $\mathcal{P}^+$ for $b = c$.

More generally, one may define projectors for any subset $S \subset \mathbb{N}^{3d}$ of monomials $p^a q^b \tau^c$. Such a subset generates a $\mathbb{C}[[\omega]]$-module M_S in the power series ring R. By the same method as above one can define the projection maps for $b \in \{c, h\}$

$$\pi^b_{S,n,m} : E^b_n \longrightarrow E^b_m, (n < m),$$

on the space $M_S \cap E^b_m$ 'spanned' by the monomials in S. The corresponding norm sequence $(|\pi^b_{S,n,n+\varepsilon}|)$ belongs to $\mathcal{P}$.

Poisson derivations. We introduce the space of Poisson derivations

$$\Theta^c_{n,m} = \Theta(R) \cap L(E^c_n, E^c_m).$$

The quantity

$$\|u\| = \sup_{n<p<q<m} \{(s_p - s_q)|u(x)|_q/(e|x|_p)\}$$

is well-defined and defines a Banach space structure on $\Theta^c_{n.m}$. (The normalisation constant $e \approx 2.718$ is purely conventional and serves to simplify the estimates of the Borel lemma 5.4.)

The map $H_{n,m} : E^c_n \longrightarrow \Theta^c_{n,m}$ $(n < m)$, which associates to a function f its hamiltonian field $\{-, f\}$ is well-defined, and the norm sequence for $H_{n,n+1}$ belongs to $\mathcal{P}$, again by Cauchy-Nagumo (Lemma 5.1).

Truncation maps for derivations factor:

$$\Theta^c_{n,n'} \xrightarrow{ev} (E^h_{n'})^{4d} \longrightarrow (E^h_{n'})^{4d} \longrightarrow \Theta^c_{m',m},\ n < n' < m' < m$$

where ev is the map which evaluates derivations on the coordinates τ, ω, q, p and the middle map of the diagram corresponds to the truncation

$$x \mapsto [x]^{2^{n+1}+1}_{2^n+1}$$

Therefore the associated norm sequence for any choice of $m = n + \varepsilon$, n', m' depending on n belongs to $\mathcal{P}^+$. We will denote this map also by $\tau_{n,m}$ when n, n', m', m are at equal mutual distances.

The map L^c. Recall that in iteration scheme for the Hamiltonian normal form of §2 used a specific $\mathbb{C}[[\omega, \tau]]$-linear map

$$L : R \longrightarrow \Theta(R) = \mathrm{Ham}\,(R) \oplus \mathrm{Der}\,(R_0), m \mapsto Lm$$

that was defined by setting for $a \neq b$:

$$Lp^a q^b := \{-, \frac{1}{(\alpha + \omega, a - b)} p^a q^b\},\ a \neq b.$$

and

$$Lg(qp) := \sum_{i=1}^{d} \frac{\partial g(\tau)}{\partial \tau_i} \partial_{\omega_i}.$$

We have to realise this on the level of Banach spaces.

We will define for $n < m' < m'' < m$ a map

$$L^c_{n,m',m'',m} : E^c_n \longrightarrow \Theta_{m',m}$$

as a composition of five basic maps.

Inclusion step:

$$i_n : E^c_n \longrightarrow E^h_n.$$

Clearly, the norm sequence belongs to $\mathcal{P}$.

Truncation step:

$$\tau_{n,m'} : E^h_n \longrightarrow E^h_{m'}.$$

As remarked before, its norm sequence belongs to $\mathcal{P}^+$.

Division step: We now use the map

$$div_{m'} \longrightarrow div_{m'},$$

where the map $div_n : E^h_n \longrightarrow E^h_n$ is defined as by

$$p^I q^J \mapsto \frac{1}{(\alpha+\omega, I-J)} p^I q^J$$

on $(M^h_n)^\perp$ and equal to the identity on M^h_n. As $\omega \in Z_n$, we have

$$|\frac{1}{(\alpha+\omega, I-J)} p^I q^J|_n \leq \frac{1}{a_n} |p^I q^J|_n.$$

As the monomials $p^I q^J$ are orthogonal, we deduce that the norm sequence of div_n has its norm bounded by $a^{-1} \in \mathcal{P}$.

Restriction step: After that, we use the restriction map

$$r_{m',m''} : E^h_{m'} \longrightarrow E^c_{m''}$$

As remarked above, the norm sequence of $r_{m+\varepsilon,m+2\varepsilon}$ belongs to $\mathcal{P}$.

Poisson step: We use the map $D_{m'',m}$, where the map

$$D_{n,m} : E^c_n \longrightarrow \Theta^c_{n,m}$$

is defined on the level of monomials by

$$\begin{cases} Dp^I q^J = \{-, p^I q^J\}, & I \neq J, \\ Dg(pq) = \sum_{i=1}^d \frac{\partial g(\tau)}{\partial \tau_i} \partial_{\omega_i}. \end{cases}$$

As we are dealing with a projection and partial differential operators, its norm is again in $\mathcal{P}$.

So we define

$$L^c_{n,m',m'',m} := D_{m'',m} \circ r_{m',m''} \circ div_{m'} \circ \tau_{m',n} \circ i_n$$

as the composition of these five maps:

$$E^c_n \xhookrightarrow{i_n} E^h_n \xrightarrow{\tau^h_{n,m'}} E^h_{m'} \xrightarrow{div_{m'}} E^h_{m'} \xrightarrow{r_{m',m''}} E^c_{m''} \xrightarrow{D_{m'',m}} \Theta^c_{m'',m}$$

We will make specific, but arbitrary choices for the intermediate points by setting

$$m' := n+\varepsilon, \;\; m'' := n+2\varepsilon, \;\; m = n+3\varepsilon, \;\; \varepsilon := (m-n)/3$$

and use the choices to define

$$L^c_{n,m} := L^c_{n,m',m'',m}$$

In our iteration we will use specifically $L^c_{n,n+1/4}$ and $L^c_{n+1/6,n+1/3}$.

Lemma 3.17. *The norm sequences* $|L^c_{n,n+1/4}|$ *and* $|L^c_{n+1/6,n+1/3}|$ *belong to* $\mathcal{P}^+$

Proof. The norm of the composition is at most the product of the norms, which are all in $\mathcal{P}$; the norm of truncation is in $\mathcal{P}^+$. Hence the result follows. □

The map j. We proceed to the definition of the map j inside our functional spaces. Recall that in §2 we defined it in term of L:

$$j_A : R \longrightarrow \Theta(R), \quad m \mapsto L(m - Lm(T)).$$

It is quite simple to realise this on the level of Banach spaces; as we need to apply it to only to specific elements

$$A_n = A_0 + T_n, \quad T_n = \sum_{i=1}^{n} S_n,$$

we consider the map

$$M_n : E^c_n \longrightarrow E^c_{n+1/4},\ m \mapsto m - L^c_{n,n+1/4}(m)T_n.$$

Note that the map

$$L^c_{n,n+1/4} : E^c_n \longrightarrow \Theta^c_{n+1/6,n+1/4},$$

so that $L^c_{n,n+1/4}(m) \in \Theta^c_{n+1/6,n+1/4}$ indeed can be applied to $T_n \in E^c_n$. We can now define a map

$$j^c_n : E^c_n \longrightarrow \Theta^c_{n+1/3,n+1/2},\ x \mapsto L^c_{n+1/4,n+1/2} \circ M_n(x)$$

that emulates the map

$$j_n : x \mapsto [j_{A_n}([x]^{2^{n+2}+2}_{2^{n+1}+2})]^{2^{n+1}}_{2^n} = [j_{A_n}([x]^{2^{n+2}+2}_{2^{n+1}+2})]^{2^{n+1}}.$$

defined in the formal KAM-iteration.

Note that by our implementation of L, we incorporated the truncation step, which leads to a norm sequence in $\mathcal{P}^+$.

Lemma 3.18. *We have* $\frac{|j^c|}{1+|T|} \in \mathcal{P}^+$.

Proof. By definition one has:

$$|j^c_n| \leq |L^c_{n,n+1/4}|\,|L^c_{n+1/4,n+1/2}|(1 + |T_n|) \in (1 + |T|)\mathcal{P}^+.$$

□

3.7. **The iteration in the Banach spaces E_n^c.** We may now formulate the Hamiltonian normal form iteration in the spaces E_n^c for appropriate choices of ρ and s_0.

Definition 3.19. *We will simplify notation slightly and write:*

$$\tau_n := \tau^c_{n+1/2,n+1} : E^c_{n+1/2} \longrightarrow E^c_{n+1}$$

for the maps that emulates the truncation $[-]_{2^n+1}^{2^{n+1}+1}$, *and similarly*

$$\sigma_n := \sigma^c_{n,n+1/2} : E^c_n \longrightarrow E^c_{n+1/2}$$

for the map the emulates the tail-truncation $[-]_{2^{n+1}+2}$ *and* j_n *for* j^c_n.

We are given an arithmetic class $\mathcal{C}(a)_\infty$ depending on a sequence $a = (a_n)$ and, without loss of generality, we may assume that $a_0 \leq 1$. We also fix an arbitrary $\alpha \in]3/2, 2[$.

We now list all the estimates which appear in the proof of the convergence. The precise form of these estimates is irrelevant for the proof of the convergence, since the absorption principle implies that these are fulfilled for sufficiently small ρ. More precisely, according to Proposition 3.16, as the norm sequences $|j|/(1+|T|)$, $|\tau|$, $|\sigma|$ belong to $\mathcal{P}^+$, we may find ρ such that for some constant $R \geq 1$ we have:

$$|j_0| \leq \frac{Ra_0(s_{1/2} - s_1)}{2}, \tag{1}$$

$$\frac{|j_n|\,|\tau_n|}{1+|T_n|} \leq \frac{Ra_n(s_{n+1/4} - s_{n+1/2})}{2e(n+1)}, \tag{2}$$

$$|\tau_n| \leq \frac{R}{2}, \tag{3}$$

$$\frac{|j_n|}{1+|T_n|} \leq \frac{Ra_n(s_{n+1/2} - s_{n+1})}{8(n+1)}, \tag{4}$$

$$|\sigma_n| \leq \frac{R^2 e^{-\alpha^n}}{8}, \tag{5}$$

$$|\sigma_n|\,|j_n| \leq \frac{R^2 e^{-\alpha^n}}{8(n+1)}. \tag{6}$$

Note that (1) follows form the $n = 0$ case of (4) and that some estimates are redundant. As A_0, B_0 vanish at the origin, we may choose s_0 small enough so that

$$\begin{array}{ll} (a) & |A_0| \leq 1, \\ (b) & |B_0| \leq R^{-1}, \\ (c) & |B_0| \leq R^{-2} e^{1/(2-\alpha)}. \end{array}$$

Clearly (c) is in general stronger than (b)! The estimates used in the proof will be numbered by letters and numbers; for instance $(1b)$ means that we use the estimates (1) and (b).

Theorem 3.20 ([14]). *Consider an analytic Hamiltonian of the form*

$$H = \sum_{i=1}^{n} \alpha_i p_i q_i + O(3) \in \mathbb{C}\{p, q\}$$

and put

$$F_0 := H + \sum_{i=1}^{d} \omega_i p_i q_i = A_0 + B_0$$

Assume that the frequency vector $\alpha \in \mathcal{C}(a)_\infty$, where a is a Bruno sequence. Under the above assumptions on the sequence ρ and s_0, the iteration

$$\begin{aligned} A_{n+1} &= A_n + S_{n+1} \\ B_{n+1} &= \phi(v_n)A_n + \psi(v_n)S_{n+1} + e^{-v_n}(\sigma_n(B_n - v_n(A_n))) \end{aligned}$$

with

$$S_{n+1} := \tau_n(B_n - v_n(A_n)), \quad v_n := j_n(B_n) = j_n B_n$$

and

$$\phi(z) = e^{-z}(1+z) - 1, \quad \psi(z) = e^{-z} - 1$$

is well-defined in the Banach space E_n^c, i.e. for all $n \in \mathbb{N}$ we have:

i) $A_n, B_n, S_n \in E_n^c$.
ii) The sequence (B_n) converges quadratically to zero.
iii) The vector fields v_n exponentiate to elements

$$\varphi_n = e^{-v_n} \in L(E_n, E_{n+1})$$

Moreover the composition

$$\Phi_n = \varphi_{n-1}\varphi_{n-2}\dots\phi_1\phi_0$$

converges to a Poisson morphism

$$\Phi_\infty \in L(E_0^c, E_\infty^c),$$

which reduces F_0 to its Hamiltonian normal form.

Remark. It could happen that the real part of the set Z_∞ reduces to the origin as the set $\mathcal{R}(a)$ might have density 0 at the point $\alpha \in \mathbb{R}^d$. In this case, our statement would simply be empty. As we already explained, this difficulty however can always be circumvented by an appropriate choice of the sequence a. According to Proposition 3.6, given a frequency vector $\alpha \in \mathbb{R}^d$ satisfying the Bruno condition, we can always find a Bruno sequence $a < \sigma(\alpha)$, for which $\mathcal{R}(a)_\infty$ has density 1 at the point α. So the above pathology can be avoided easily by an appropriate choice of the Bruno sequence $a = (a_n)$. This is a fundamental difference between KAM theory and Diophantine analysis, where one is concerned with optimal bounds [15, 19, 20].

Proof. We start with

$$F_0 = A_0 + B_0,$$

and consider A_0 and B_0 as elements of E_0^c. We have:

$$v_0 = j_0(B_0) \in \Theta^c_{1/4,1/2}$$

and, as $T_0 = 0$, we get the estimate:

$$(1b) \qquad |v_0| \leq |j_0||B_0| \leq \frac{a_0(s_{1/2} - s_1)}{2}.$$

According to the Borel lemma (Lemma 5.4), the linear map

$$e^{-v_0} : E^c_{1/2} \longrightarrow E^c_1$$

is well-defined and moreover, as e^z is the Borel transform of $1/(1-z)$, we deduce from the previous estimate that:

$$|e^{-v_0}| \leq \frac{1}{1 - 1/2} = 2.$$

In the next step we get

$$F_1 = e^{-v_0} F_0 \in E_1^c,$$

and $A_1 = A_0$, $S_1 = 0$.

We show by induction the following estimates:

$$\begin{aligned} &|S_{n+1}| \leq R|B_n|, \\ &|B_{n+1}| \leq \frac{R^2}{2}|B_n|^2 + \frac{R^2}{2}e^{-\alpha^n}|B_n|, \\ &|B_{n+1}| < R^{-1}. \end{aligned}$$

Assuming the validity of these estimates up to index n, we may conclude that

$$(b) \qquad |T_n| = |\sum_{k=1}^n S_k| \leq \sum_{k=1}^n |S_k| \leq R \sum_{k=1}^n |B_k| \leq n.$$

In particular, we may simplify the estimates in which $1+|T_n|$ is involved. Note also that:

$$(a) \qquad |A_n| \leq |A_0| + |T_n| \leq n + 1.$$

As formal power series, the term S_{n+1} is the difference of two terms

$$S_{n+1} = \tau_n(B_n - j_n(B_n)(A_n)) = \tau_n(B_n) - \tau_n(j_n(B_n)(A_n)).$$

Inside our Banach spaces, the second term

$$J_n(B_n) := \tau_n(j_n(B_n)(A_n))$$

can be estimated as follows. The map J_n is obtained as composition

$$E_n^c \xrightarrow{j_n} \Theta^c_{n+1/4,n+1/2} \xrightarrow{ev_n} E^c_{n+1/2} \xrightarrow{\tau_n} E^c_{n+1},$$

where the evaluation map ev_n is defined by

$$ev_n(v) = v(A_n).$$

Note that by definition of the norm in $\Theta^c_{n+1/4,n+1/2}$, we have:

$$|v(A_n)| \leq \frac{e|A_n|}{a_n(s_{n+1/4} - s_{n+1/2})}.$$

Consequently:

$$|J_n| \leq |j_n|\,|\tau_n| \frac{e|A_n|}{a_n(s_{n+1/4} - s_{n+1/2})} \leq \frac{R}{2}, \tag{2}$$

$$|S_{n+1}| \leq (|\tau_n| + |J_n|)|B_n| \leq R|B_n|. \tag{3}$$

This proves the first step of the induction.

We can now form

$$A_{n+1} = A_n + S_{n+1} \in E^c_{n+1},$$

and set as usual:

$$v_n := j_n(B_n).$$

Now:

$$|v_n| \leq |j_n||B_n| \leq \frac{a_n(s_{n+1/2} - s_{n+1})}{8} < \frac{a_n(s_{n+1/2} - s_{n+1})}{2}, \tag{4b}$$

and therefore by the Borel lemma (Lemma 5.4), the linear map

$$e^{-v_n} : E^c_{n+1/2} \longrightarrow E^c_{n+1}$$

is well-defined and, as e^{-z} is the Borel transform of $1/(1+z)$, we also get that:

$$|e^{-v_n}| \leq \frac{1}{1-1/2} = 2.$$

So we can form $F_{n+1} = e^{-v_n} F_n \in E^c_{n+1}$ and so indeed

$$B_{n+1} = F_{n+1} - A_{n+1} \in E^c_{n+1}.$$

Let us now prove the announced central estimate of $|B_{n+1}|$. The power-series

$$e^{-z}(1+z) - 1 \in \mathbb{C}\{z\}$$

is the Borel transform of

$$-\frac{z^2}{(1+z)^2} \in z^2\mathbb{C}\{z\},$$

which has radius of convergence equal to 1 and, choosing $r \leq 1/2$, we get that:

$$\left|\frac{1}{(1+z)^2}\right| \leq \frac{1}{(1-r)^2} \leq 4.$$

We apply the Borel lemma (Lemma 5.4) and obtain:

$$|\phi(j_n B_n)| \leq 4\left(\frac{|j_n|\,|B_n|}{a_n(s_{n+1/2}-s_{n+1})}\right)^2 \leq \frac{R^2|B_n|^2}{4}. \tag{4}$$

Similarly the series $\psi(z)$ is the Borel transform of $z/(1-z)$ and, as

$$|S_{n+1}| < |B_n| \text{ and } \frac{1}{1-1/2} = 2,$$

the Borel estimate gives

$$|\psi(j_n(B_n))S_{n+1}| \leq 2\frac{|j_n|\,|B_n|\,|S_{n+1}|}{a_n(s_{n+1/2}-s_{n+1})} \leq \frac{R^2|B_n|^2}{4}. \tag{4}$$

Finally we look at the remainder term $e^{-v_n}(\sigma_n(B_n - j_n(B_n)(A_n))$:

$$|e^{-v_n}\sigma_n(B_n)| \leq 2|\sigma_n||B_n| \leq \frac{1}{4}e^{-\alpha^n}R^2|B_n|, \tag{5}$$

$$|e^{-v_n}(\sigma_n(j_n(B_n)(A_n)) \leq 2|\sigma_n|\,|j_n|\,|A_n|\,|B_n| \leq \frac{R^2e^{-\alpha^n}}{4}|B_n|. \tag{6}$$

This proves the second step of the induction

$$|B_{n+1}| \leq \frac{R^2}{2}|B_n|^2 + \frac{R^2}{2}e^{-\alpha^n}|B_n|.$$

The 'norm map':

$$E^c := \bigsqcup_n E_n^c \longrightarrow \mathbb{R}_+,\ \eta \mapsto |\eta|$$

reduces the issue of convergence to the analysis of some elementary iterations of positive numbers.

So let us put $x_n := |B_n|$ so that

$$x_{n+1} \leq \frac{R^2}{2}x_n^2 + \frac{R^2}{2}e^{-\alpha^n}x_n,$$

where we have $x_0 \leq R^{-2}e^{-1/(2-\alpha)}$.

We show that (x_n) converges quadratically to zero. To see this, consider the real sequence (y_n) defined by

$$y_0 = R^{-2}e^{-1/(2-\alpha)},\ y_{n+1} = \frac{R^2}{2}\left(e^{\alpha^n}y_n^2 + e^{-\alpha^n}y_n\right),$$

which clearly majorates the sequence (x_n). It follows with an easy induction that one has the inequality

$$y_n \geq e^{-2\alpha^n}.$$

Indeed, assuming the truth for y_n, we get

$$y_{n+1} \geq \frac{R^2}{2}\left(e^{-3\alpha^n} + e^{-3\alpha^n}\right) = R^2 e^{-3\alpha^n} \geq e^{-2\alpha^{n+1}},$$

as $3/2 < \alpha < 2$ and $R > 1$. The inequality can also be written as:

$$e^{-\alpha^n} y_n < e^{\alpha^n} y_n^2,$$

and therefore:

$$y_{n+1} = \frac{R^2}{2}\left(e^{\alpha^n} y_n^2 + e^{-\alpha^n} y_n\right) < \frac{R^2}{2}\left(e^{\alpha^n} y_n^2 + e^{\alpha^n} y_n^2\right) = R^2 e^{\alpha^n} y_n^2.$$

This shows that the sequence

$$z_0 = R^{-2} e^{-1/(2-\alpha)},\ z_{n+1} = R^2 e^{\alpha^n} z_n^2$$

majorates both (y_n) and (x_n). This sequence is easily integrated

$$z_n = R^{2^{n+1}-2} e^{\beta_n} z_0^{2^n}.$$

with

$$\beta_n := 2^{n-1} \sum_{k=0}^{n-1} \left(\frac{\alpha}{2}\right)^k = 2^{n-1} \frac{1-(\alpha/2)^n}{1-(\alpha/2)} \sim 2^n \frac{1}{2-\alpha}.$$

Writing (z_n) in the form

$$z_n = R^{-2}(R^2 e^{\gamma_n} z_0)^{2^n},\ \gamma_n := \frac{1-(\alpha/2)^n}{2-\alpha}.$$

The sequence γ_n increases and is bounded by $1/(2-\alpha)$ thus:

$$(c) \qquad R^2 e^{\gamma_n} z_0 \leq R^2 e^{1/(2-\alpha)} z_0 < 1,$$

thus the sequence (z_n) is decreasing and converges quadratically to zero. This shows the quadratic convergence of (B_n) and concludes the proof of (b). The quadratic convergence for $v_n = j_n(B_n)$ follows from this and the Composition Lemma (Lemma 5.3) implies (c). □

3.8. **Regularity of the normal form.** We formulated the iteration scheme in terms of the sequence of Banach spaces E_n^c. Without much difficulty one can formulate a version of the iteration in the spaces E_n^k. Without going into all details, we state the following

Proposition 3.21 ([14]). *Under the assumptions of Theorem 3.20, the iteration is well-defined in E_n^k and the sequences (B_n^k) and (v_n^k) converge quadratically to zero. Furthermore, the Poisson morphism Φ_∞ maps $E_0^k \subset E_0^c$ to $E_\infty^k \subset E_0^c$.*

Proof. This is a direct consequence of Pöschel's regularity lemma (5.8). Indeed, the norms $\nu = (\nu_n)$ of the maps

$$E_n^c \longrightarrow E_{n+1}^k$$

are bounded by a positive Bruno sequence. But the sequences of the iteration converge to zero quadratically, so the multiplication by ν has no effect on the convergence. □

We remark that the Local Equivalence Lemma (Lemma 5.2) implies that we can also formulate and prove quadratic convergence for the iteration in (E_n^h).

4. Application to invariant tori

We describe now an application of our theorem to the analysis of invariant tori near elliptic critical points of analytic Hamiltonians, whose frequency satisfies a Bruno condition. For this we have to consider the appropriate real form of H and restrict to the real domain.

4.1. **Hyperbolic and Elliptic fixed points.**

The dynamics of the harmonic oscillator

$$H_e = \frac{1}{2}\sum_{i=1}^{d} \beta_i(p_i^2 + q_i^2)$$

describes quasi-periodic motions with frequency vector β. All orbits are bounded and the phase space is filled out by a d-parameter family of invariant tori $p_i^2 + q_i^2 = t_i$, on which the solutions spiral around. The geometry of the situation is well-known: the fibres of the map

$$\mathbb{R}^{2d} \longrightarrow \mathbb{R}^d_{>0}, \quad (q,p) \mapsto p^2 + q^2 := (p_1^2 + q_1^2, \ldots, p_d^2 + q_d^2)$$

are tori, which are of real dimension d over the strictly positive orthant $\mathbb{R}^d_{>0}$.

In the real domain there is a big difference in the dynamical behaviour between H_e and its hyperbolic cousin

$$H_h = \sum_{i=1}^{d} \alpha_i p_i q_i,$$

for which all orbits are unbounded and there exist no invariant tori.

Yet when considered over $\mathbb{C}$, the canonical coordinate transformation ϕ

$$p_j \mapsto \frac{1}{\sqrt{2}}(p_j + iq_i), \quad q_j \mapsto \frac{1}{\sqrt{2}}(q_j + ip_j)$$

maps H_h to H_e, when we put

$$\beta = i\alpha.$$

Another way of expressing the relation between H_h and H_e is by saying that the evolution for H_h in purely imaginary time is equivalent to the real time evolution of H_e and vice versa. As a consequence of this relation, we can immediately translate results about H_h into results about H_e.

4.2. **The coordinate transformation.** Consider an analytic hamiltonian of the form

$$H = \frac{1}{2}\sum_{i=1}^{d} \alpha_i (p_i^2 + q_i^2) + O(3) \in \mathbb{R}\{p, q\}.$$

and assume that the frequency vector $\alpha \in \mathcal{R}(a)_\infty$, where a is a Bruno sequence. We can apply our theorem 3.20, so for appropriate choice of the sequence ρ and radius s_0, we find sets

$$W_0 = Z_0 \times D_0 \times D_0^2, \quad Z_0 = B(s_0), \quad D_0 := D_{s_0}^d$$

and

$$W_\infty = Z_\infty \times D_\infty \times D_\infty^2, \quad D_\infty := D_{s_\infty}^d,$$

such the sequence

$$A_0, \quad A_1, \quad A_2, \dots$$

which converges in the Banach space $\mathcal{O}^c(W_\infty)$ to an element

$$A_\infty \in \mathcal{O}^c(W_\infty)$$

and the sequence

$$\Phi_0 = e^{-v_0}, \quad \Phi_1 = e^{-v_1}e^{-v_0}, \ \dots, \Phi_n = \prod_{i=0}^{n} e^{-v_k},$$

converges in the operator norm to $\Phi_\infty \in L(\mathcal{O}^c(W_0), \mathcal{O}^c(W_\infty))$. This transformation maps, for any k, the subspace $\mathcal{O}^k(W_0)$ to $\mathcal{O}^k(W_\infty)$ (Proposition 3.21). In particular, if the closed set W_0 is chosen inside the holomorphy domain of A_0, then it is in particular C^∞ on W_0 and therefore belongs to $\mathcal{O}^k(W_0)$ for any k. The function A_∞ is then C^∞ on W_∞ and for fixed $\omega \in Z_\infty$ it is holomorphic.

Of course, in a sense we get a 'half-way theorem', as we start with a real Hamiltonian, but obtain a statement about its behaviour in the complexified domain. But it is clear from the explicit form of the description of the iteration that, starting from a real Hamiltonian, the algorithm produces *real* vector fields v_n, which exponentiate to *real* analytic coordinate transformations $\varphi_n = e^{-v_n}$, etc. As a consequence the limit transformation Φ_∞ is 'real'. Furthermore, we remark that it follows from the construction of the vector fields v_n that the transformation φ_n maps the subspace $F_n^c = \mathcal{O}^c(V_n)$, $(V_n = W_n \times D_n)$ to F_{n+1}^c, so that Φ_∞ maps F_0^c to F_∞^c and by the regularity property F_0^k to F_∞^k.

The coordinate functions ω, τ, q, p can be considered as elements of the space $\mathcal{O}^c(W_0)$ and we write

$$\omega' = \Phi_\infty(\omega), \quad \tau' = \Phi_\infty(\tau), \quad q' = \Phi_\infty(q), \quad p' = \Phi_\infty(p).$$

Note that $\tau' = \tau$ and

$$\omega' \in F_\infty^k = \mathcal{O}^k(V_\infty), \quad \text{for all } \ k \in \mathbb{N},$$

so is independent of q, p. These functions define a C^∞-map

$$\phi' : W_\infty \longrightarrow W_0, \quad x \mapsto (\omega'(x), \tau'(x), q'(x), p'(x)), \ x = (\omega, \tau, q, p)$$

and for each $g \in \mathcal{O}^c(W_0)$ we have the relation

$$g(\phi'(x)) = \Phi_\infty(g)(x).$$

The reality of Φ_∞ implies that the map ϕ' maps the real part

$$\mathcal{W}_\infty := W_\infty \cap \mathbb{R}^d \text{ to } \mathcal{W}_0 := W_0 \cap \mathbb{R}^d.$$

Thus we obtain a real C^∞-map

$$\varphi' : \mathcal{W}_\infty \longrightarrow \mathcal{W}_0.$$

As the φ' sends the (ω, τ)-space to itself, the map is fibred over the (ω, τ)-space and we obtain a commutative diagram:

$$\begin{tikzcd} \mathcal{W}_\infty \arrow[r, "\varphi'"] \arrow[d] & \mathcal{W}_0 \arrow[d] \\ \mathcal{V}_\infty \arrow[r, "\psi'"] & \mathcal{V}_0 \end{tikzcd}$$

with $\psi' = (\omega', \tau')$ and vertical maps in the diagram forget the coordinates q, p.

Recall the following Whitney extension theorem:

Theorem 4.1 ([38])**.** *Let $X \subset \mathbb{R}^d$ be a compact subset. Any function $f \in C^\infty(X, \mathbb{R})$ is the restriction of a C^∞ function defined on $\mathbb{R}^d$.*

Invoking this theorem to the component functions ω', τ', q', p' of φ', we obtain a C^∞-maps

$$\psi : \mathbb{R}^{2d} \longrightarrow \mathbb{R}^{2d}, \quad \varphi : \mathbb{R}^{4d} \longrightarrow \mathbb{R}^{4d}.$$

We restrict ψ and φ to the preimages

$$\mathcal{V}_e := \psi^{-1}(\mathcal{V}_0) \supset \mathcal{V}_\infty, \quad \mathcal{W}_e := \varphi^{-1}(\mathcal{W}_0) \supset \mathcal{W}_\infty,$$

and we arrive at a diagram

$$\begin{tikzcd} \mathcal{W}_e \arrow[r, "\varphi"] \arrow[d] & \mathcal{W}_0 \arrow[d] \\ \mathcal{V}_e \arrow[r, "\psi"] & \mathcal{V}_0 \end{tikzcd}$$

that extends the previous diagram. Obviously, the maps φ and ψ are not unique, but its restrictions to $\mathcal{W}_\infty$ and $\mathcal{V}_\infty$ are.

As the Taylor series of φ is given by the series Φ_∞, which is $Id + O(2)$, φ is a diffeomorphism near the origin. Consequently, by restriction to smaller polydiscs, we may and will assume that

(1) ψ is a diffeomorphism between $\mathcal{V}_e$ and $\mathcal{V}_0$,
(2) φ is a diffeomorphism between $\mathcal{W}_e$ and $\mathcal{W}_0$.

As the map φ arose from the transformation Φ_∞, it has the property that after restriction to $\mathcal{W}_\infty$, it transforms

$$F_0 = A_0 + H = \frac{1}{2}\sum_{i=1}^{d} \omega_i(p_i^2 + q_i^2) + H(p,q)$$

to $F_\infty = A_\infty$. Furthermore

$$A_\infty = A_0 + T_\infty, \quad T_\infty \in (R_0 + I^2) \cap \mathcal{O}^k(W_\infty).$$

This means that for $\omega \in Z_\infty$ one has

$$F_0 \circ \varphi(\omega, \tau, q, p) = A_\infty(\omega, \tau, q, p),$$

and moreover, for such a value, the map $\varphi(\omega, -)$ is an analytic Poisson morphism in the variables (τ, q, p).

4.3. **Frequency maps.** In this section we return to the complex situation and start to analyse the limit $n \longrightarrow \infty$ of the frequency manifolds X_n that were considered in section 2. We need a more careful use of the Whitney extension theorem.

We consider the extension diagram of neighbourhoods of the previous section, but in the complex setting:

$$\begin{tikzcd} W_\infty \arrow[r] \arrow[d] & W_e \arrow[r, "\phi"] \arrow[d] & W_0 \arrow[d] \\ V_\infty \arrow[r] & V_e \arrow[r, "\Psi"] & V_0 \end{tikzcd}$$

We consider

$$X_n = \{(\omega, \tau) \in V_n \mid \ R_{n,1}(\omega, \tau) = \cdots = R_{n,d}(\omega, \tau) = 0\},$$

where $R_{n,i} := \Phi_n(\omega_i) \in \mathcal{O}^k(V_n)$. We will have to take the degree of differentiability $k \geq 1$. For $n = \infty$ we have the functions

$$R_{\infty,i} = \Phi_\infty(\omega_i) \in F_\infty^k = \mathcal{O}^k(V_\infty),$$

which are analytic in τ and C^k in ω.

We can consider the *limit set*

$$X_\infty := \{(\omega, \tau) \in V_\infty \mid \ R_{\infty,1}(\omega, \tau) = \cdots = R_{\infty,d}(\omega, \tau) = 0\}$$

and invoke the following theorem due to Fefferman :

Theorem 4.2 ([7]). *Let $X \subset \mathbb{C}^N$ be a compact subset and $k \in \mathbb{N}$. There exists a bounded linear operator*

$$T : C^k(X, \mathbb{C}) \longrightarrow C^k(\mathbb{C}^N, \mathbb{C}),$$

which is right inverse to the restriction mapping.

We consider the inclusion

$$V_\infty \subset V_e$$

with $k = 1$ and get an extension operator T. We consider the functions

$$R'_{n,i} := R_{n,i}|V_\infty, \ n \in \mathbb{N}$$

and set

$$r_{n,i} := T(R'_{n,i}), \ r_n := (r_{n,1}, \dots, \ r_{n,d}).$$

As T is a bounded operator, we can conclude the existence of a limit

$$r_\infty = \lim_{n \longrightarrow \infty} r_n$$

By the implicit function theorem each manifold

$$X_{n,e} := \{(\omega, \tau) \in V_e : r_n(\omega, \tau) = 0\}, \ n \in \mathbb{N} \cup \{\infty\}$$

is, near the origin, the graph of a function

$$\{\omega = f_n(\tau)\}.$$

These functions can be constructed using Picard iteration, whose convergence is controlled by a condition on the derivative dr_n. Since (r_n) converge in the C^1-topology, we may find a *common compact neighbourhood* of the origin for all these maps

$$f_n : U \longrightarrow U'.$$

So by possible shrinking of V_e to the set $U \times U' \subset \mathbb{C}^d \times \mathbb{C}^d$ we can express the manifolds $X_{n,e}$ has graphs over a common neighbourhood.

Note that these functions converge pointwise to f_∞. Indeed, take $\tau \in U$, as U' is compact $(f_n(\tau))$ admits converging subsequences. Let ω be a limit value of such a subsequence $(f_{n_k}(\tau))$. Passing to the limit in the equality

$$r_{n_k}(f_{n_k}(\tau), \tau) = 0$$

we get that the point (τ, ω) belongs to the manifold $X_{\infty,e}$ and therefore $\omega = f_\infty(\tau)$. We will not use this fact.

The map

$$U \longrightarrow \mathbb{C}^d, \ \ \tau \mapsto \alpha + f_\infty(\tau).$$

will called a *frequency map* of our Hamiltonian system.

According to 2.5, the Taylor series of this map at the origin at order k coincides with that of the formal frequency map b. Of course, the process of extension is not unique, but the restriction to the preimage of $\mathcal{C}(a)_\infty$ is the same for any choice. The construction can done so that the resulting frequency map restricts to a real map

$$\beta : \mathcal{U} \longrightarrow \mathbb{R}^d, \ \ \tau \mapsto \alpha + f_\infty(\tau).$$

4.4. **The non-degeneracy condition.** We now wish to construct a frequency map whose image lies inside $\alpha + F(H)$. Later, this condition will guarantee positive measure for the set of invariant tori. This property depends on our choice of Whitney extension, but there is an intrinsic underlying statement: Proposition 2.9 stating that all X_n's are contained in $F(H) \times \mathbb{C}^d$ extends to $n = \infty$.

Proposition 4.3. *The limit set X_∞ satisfies the same non-degeneracy condition as the manifolds X_n:*

$$X_\infty \subset F(H) \times \mathbb{C}^d.$$

Proof. Consider a unit vector $n \in \mathbb{C}^d$ normal to the frequency space $F(H)$ induces, via the Euclidean scalar product, a linear form

$$u : \mathbb{C}^{2d} \longrightarrow \mathbb{C}, x \mapsto (x, n),\ F(H) \subset \operatorname{Ker} u$$

Assume that there exists a point $x_0 = (\omega_0, \tau_0) \in X_\infty$ which does not lie inside the frequency space:

$$u(x_0) \neq 0 \text{ and } R_\infty(x_0) = 0.$$

The point $x_0 = (\omega_0, \tau_0)$ remains at distance $K = |u(x_0)|$ from the hyperplane $\operatorname{Ker} u$. By Proposition 2.9, the manifolds X_n are contained inside the frequency space therefore the point (ω, τ) remains also at distance K from all the manifolds

$$X_n = \{(\omega, \tau) \in V_e : R_n(\omega, \tau) = 0\}.$$

We consider the Whitney extensions:

$$X_{n,e} = \{(\omega, \tau) \in V_e : r_n(\omega, \tau) = 0\}$$

and as both manifolds intersected with $\mathcal{C}(a) \times \mathbb{C}^d$ are the same the points of

$$(X_{n,e} \cap \{\omega = \omega_0\}) = (X_n \cap \{\omega = \omega_0\})$$

are also at distance at least K from the point $x_0 \in X_\infty$.

Now let the parameter τ vary and define the functions

$$h_{n,i}(\tau) = r_{n,i}(\omega_0, \tau),\ n \in \mathbb{N} \cup \{+\infty\}.$$

By definition of our functional spaces, as $\omega_0 \in \mathcal{C}(a)$, these functions are holomorphic in τ even for $n = +\infty$.

At least one of the functions $h_{\infty,i}$ is not constant for some $i = 1, \dots, d$, otherwise this common constant will necessarily be zero and x_0 will be the origin, contradicting the assumption $x_0 \notin F(H) \times \mathbb{C}^d$.

Now choose a complex line $L \subset \mathbb{C}^d$ containing x_0 along which the function $h_{\infty,i}$ is not constant. By restriction to this line we get a sequence of one variable holomorphic functions

$$g_n : L \cap V_e \longrightarrow \mathbb{C}, z \mapsto h_{n,i}(z)$$

which has the following properties:

1) $g_\infty(x_0) = 0$.
2) $\|x - x_0\| < K \implies g_n(x) \neq 0$ for any $n \in \mathbb{N}$.
3) the sequence (g_n) converges to g_∞ in the C^1-norm.

Condition 2) 3) are in obvious contradiction with 1). Indeed The number of zeroes $Z_{n,i}$ counted with multiplicities contained inside the disk $D \subset L$ centred at x_0 of radius $K/2$ is given by the integral formula

$$Z_{n,i} = \frac{1}{2i\pi} \int_{\partial D} \frac{g'_{n,i}(z)}{g_{n,i}(z)} dz$$

for any $n \in \mathbb{N} \cup \{\infty\}$. We know that for $n \in \mathbb{N}$, $Z_{n,i} = 0$, thus passing to the limit gives $Z_{\infty,i} = 0$. This contradicts 1) and concludes the proof of the proposition. □

As we know that X_∞ is contained inside $F(H) \times \mathbb{C}^d$, we can improve our previous construction of a frequency map. First we restrict the functions $R_{\infty,i}$ to

$$V_\infty \cap \left(F(H) \times \mathbb{C}^d\right)$$

Now we use Whitney C^∞-extension to and obtain functions

$$s_{\infty,1}, \dots, s_{\infty,d} \in C^\infty(\mathbb{C}^{2d}, \mathbb{C})$$

whose common zero set, when restricted to $F(H) \times U$, can be written as a graph of a map:

$$b : U \longrightarrow \alpha + F(H) \subset \mathbb{C}^d,$$

and similarly in the real case

$$\beta : \mathcal{U} \longrightarrow \alpha + \mathcal{F}(H) \subset \mathbb{R}^d.$$

In this way, we construct frequency maps b satisfy the non degeneracy condition which will be needed to prove positive measure of invariant tori: *the partial derivatives of the frequency map evaluated at the origin generate the frequency space.* Moreover as the condition is open, we may assume, up to a possible shrinking of $\mathcal{U}$, that the condition holds not only at the origin, but at any point of $\mathcal{U}$.

4.5. **The elliptic normal form Theorem.** Using the coordinate transformation and the frequency map, we may sum up our results in the following way:

Theorem 4.4 ([11, 12])**.** *Let $a = (a_n)$ be a sequence satisfying the Bruno condition and $\alpha \in \mathcal{R}(a)_\infty$. Let $H \in \mathbb{R}\{q, p\}$ be a real analytic function with an elliptic fixed point:*

$$H = \frac{1}{2} \sum_{i=1}^{d} \alpha_i (p_i^2 + q_i^2) + O(3).$$

Then there exists an open neighbourhood of the origin $U \subset \mathbb{C}^d$, $V \subset \mathbb{C}^{2d}$ with real parts $\mathcal{U}, \mathcal{V}$ and C^∞-maps

$$\beta : \mathcal{U} \longrightarrow \alpha + \mathcal{F}(H) \subset \mathbb{R}^d,$$

$$\begin{tikzcd} \mathcal{U} \times \mathcal{V} \arrow[rr, "\Psi"] \arrow[dr] & & \mathcal{U} \times \mathbb{R}^{2d} \arrow[dl] \\ & \mathcal{U} & \end{tikzcd}$$

such that for any $\tau \in \beta^{-1}(\mathcal{R}(a)_\infty)$, one has:

i) *The Taylor series expansion of β at the origin is equal to $\nabla B(H)$.*
ii) *The map Ψ is a fibred diffeomorphism over its image.*
iii) *The map $\Psi(\tau, -)$ is an analytic symplectomorphism.*
iv) $H \circ \Psi(\tau, q, p) = \frac{1}{2}\sum_{i=1}^n \beta_i(\tau)(p_i^2 + q_i^2) + T_\infty(\tau, q, p)$
v) $T_\infty(\tau, -) \in I^2 + \mathbb{C}$, *where $I \subset \mathcal{O}^c(V)$ is the ideal generated by the $p_i^2 + q_i^2 - \tau_i$'s.*

The map Ψ of the theorem is defined in terms of the map φ and the map $\tau \mapsto \omega(\tau)$ of the previous section by the relation

$$\Psi(\tau, q, p) = \varphi(\omega(\tau), \tau, q, p).$$

We note that in the extremal case where $\mathcal{F}(H) = \{0\}$, the frequency map β is constant, and the condition $\beta(\tau) \in \mathcal{R}(a)_\infty$ is always satisfied. In this case the map Ψ is therefore analytic, because φ is analytic in the τ-variables. So the theorem implies that H is integrable, and thus we recover a classical result of Rüßmann [32].

In the general case, our iteration produces a C^∞ function β, whose Taylor expansion at the origin is the formal frequency map given by the Birkhoff normal form. In a similar way, our construction shows that the sequence (h_n) of 2.5 converges to a limit h_∞. This limit function being the constant term in the expression

$$T_\infty(\tau, q, p) = h_\infty(\tau) + \sum_{i,j} t_{ij}(q, p) f_i f_j$$

with $\tau \in \beta^{-1}(\mathcal{R}(a)_\infty)$.

The Taylor expansion of h_∞ at the origin is the Birkhoff normal form and the map β can then be chosen to be the gradient of h_∞. In analogy with Pöschel terminology [30], we might say that the situation is similar to that of the Birkhoff normal form but over a Cantor set. But we will not use this fact in the sequel.

Eliasson posed the question whether the frequency map β is analytic or not [4, 29]. We do not have an answer, but we remark that the frequency map β is constructed out of the frequency manifolds X_n and more precisely by the Malgrange-Mather division theorem [22, 23]:

$$R_\infty(\omega, \tau) = A(\omega, \tau)(\omega - \beta(\tau)).$$

We only know that the function R_∞ is analytic in τ and Eliasson's question concerns the map β. So the mystery remains...

4.6. **A big set of invariant tori.** A direct corollary of the elliptic normal form theorem is the following.

Corollary 4.5. *For $\tau \in \beta^{-1}(\mathcal{R}(a)_\infty)$, the image under $\Psi(\tau, -)$ of the torus*

$$T_\tau : p_1^2 + q_1^2 = \tau_1, \dots, p_1^2 + q_1^2 = \tau_n$$

is invariant under the Hamiltonian flow of H. The motion on this torus is quasi-periodic with frequency $\beta(\tau)$.

So we get a collection of invariant tori in our hamiltonian system, parametrised by the *inverse image* of $\mathcal{R}(a)_\infty$ by the frequency map

$$\beta : \mathcal{U} \longrightarrow \alpha + \mathcal{F}(H) \subset \mathbb{R}^d.$$

As pointed out in remark 3.7, we may and will suppose that a is chosen so that the set $\mathcal{R}(a)_\infty$ is α-dense. But without further precautions, the inverse image under β might still be a very small set, maybe reduced to an half line. We know however that our frequency map β is non-degenerate in the frequency space $\alpha + \mathcal{F}(H)$. The following *arithmetic density theorem* then can be used to control the density of the inverse image of $\mathcal{R}(a)_\infty$ with $n = d$.

Theorem 4.6 ([15])**.** *Consider a real positive decreasing sequence $\sigma = (\sigma_k)$ and let $\nu = (\nu_k)$ be a real positive sequence such that the sequence*

$$(2^{kn} \nu_k^{1/dl})$$

is summable and $\nu_k < 1$ for all k's. Consider a mapping

$$f = (f_1, \dots, f_d) : \mathbb{R}^d \supset U \longrightarrow \mathbb{R}^n$$

such that $f(U)$ is contained in an affine space spanned by the partial derivatives of f up to order l. Then the density of the set $f^{-1}(\mathcal{R}(\nu\sigma)_\infty)$ at the origin is equal to 1.

So the logic of our argument is the following: we fix a vector $\alpha \in \mathbb{R}^d$ and assume that $\sigma(\alpha) \in \mathcal{B}^-$. We consider the sequences

$$\nu_k = (2^{-(k+1)d^2 l}), \; a = \nu\sigma(\alpha).$$

Note that

$$\nu \in \mathcal{B}^- \text{ and } \sigma(\alpha) \in \mathcal{B}^- \implies a = v\sigma(\alpha) \in \mathcal{B}^-.$$

Now the elliptic normal form theorem 4.4 applies. By Proposition 4.3, the frequency mapping β we construct satisfies the assumption of the arithmetic density theorem. Consequently the set $\beta^{-1}(\mathcal{R}(a)_\infty)$ has density one at the origin and is, in particular, a set of positive measure.

We will now see how these tori fit together in a neighbourhood of the origin of our original Hamiltonian $H(p,q)$. The map

$$\Psi : \mathcal{U} \times \mathcal{V} \longrightarrow \mathcal{U} \times \mathbb{R}^{2d}$$

of the previous theorem has an inverse $\Gamma = \Psi^{-1}$ over a sufficiently small neighbourhood of the origin of the form $\mathcal{U} \times \mathcal{B}$, $\mathcal{B} \subset \mathbb{R}^{2d}$:

$$\Gamma : \mathcal{U} \times \mathcal{B} \longrightarrow \mathcal{U} \times \mathcal{V}; (\tau, p, q) \mapsto (\tau, P(\tau, p, q), Q(\tau, p, q))$$

So we have the relation

$$H(p,q) = A_0(\omega(\tau), P(\tau, p, q), Q(\tau, p, q)) + T_\infty \circ \Gamma(\tau, p, q).$$

We can, in principle, eliminate the variables $\tau_1, \tau_2, \dots, \tau_d$ from the right hand side by solving the implicit equations

$$P_i(\tau, p, q)^2 + Q_i(\tau, p, q)^2 = \tau_i, \ \ i = 1, 2, \dots, d,$$

which produces a map

$$T : \mathcal{B} \longrightarrow \mathcal{U}; \ \ (p, q) \mapsto (\tau_1(p, q), \dots, \tau_d(p, q)).$$

As one has

$$\tau_i(p, q) = p_i^2 + q_i^2 + O(3),$$

the map T is generically a submersion. In fact, it is a submersion on $\mathcal{B} \setminus C$, $C := T^{-1}(\Delta)$, where $\Delta \subset \mathcal{U}$ is the set of *critical values* of T.

One now obtains a diagram

$$\begin{tikzcd}
\mathcal{B} \arrow[r, "\gamma"] \arrow[d, "T"'] & (\alpha + \mathcal{F}(H)) \times \mathcal{V} \arrow[d, "\pi"] \\
\mathcal{U} \arrow[r, "\beta"] & \alpha + \mathcal{F}(H) \\
S \arrow[u, hook] \arrow[r] & (\alpha + \mathcal{F}(H)) \cap \mathcal{R}(a)_\infty \arrow[u, hook]
\end{tikzcd}$$

related to our normal forms as follows.

On the right hand side we have the standard Hamiltonian

$$A_0 = \frac{1}{2} \sum_{i=1}^{d} (\alpha_i + \omega_i)(p_i^2 + q_i^2),$$

defined on $(\alpha + \mathcal{F}(H)) \times \mathcal{V}$, where the map

$$\pi : (\omega, q, p) \mapsto \alpha + \omega$$

gives the frequency of motion.

On the left hand side we have a neighbourhood $\mathcal{B}$, on which the original Hamiltonian $H(p,q)$ is defined. The vertical map on the left is the τ-map $T(p,q)$ defined above.

The horizontal map γ stems from coordinate transformation Γ:

$$\gamma : \mathcal{B} \longrightarrow (\alpha + \mathcal{F}(H)) \times \mathcal{V}$$
$$(q,p) \mapsto (\beta(T(p,q)), P(T(p,q),p,q), Q(T(p,q),p,q))$$

The horizontal map in the middle is the frequency map β, which is non-degenerate in the affine sub-space $\alpha + \mathcal{F}(H)$. The inverse image $S := \beta^{-1}(\mathcal{R}(a)_\infty)$ under β parametrises invariant tori for H in the neighbourhood $\mathcal{B}$. As T is a submersion outside Δ, which by Sard's theorem has measure zero, the set $T^{-1}(S \setminus \Delta) \subset \mathcal{B}$ yields a set of positive measure consisting of invariant tori in the neighbourhood $\mathcal{B}$ of elliptic critical point, as conjectured by Herman [18].

5. Appendix: Lemmas from the hall of fame

In this appendix we collect some fundamental lemmas of great use in the type of analysis we are pursuing here. We include proofs, as these are usually elementary.

Given two open sets $V \subset U \subset \mathbb{C}^n$, we denote by $r = d(U,V)$ the supremum of the real numbers ρ for which

$$V + \rho D \subset U.$$

where D denotes the unit polydisc.

5.1. **Cauchy-Nagumo lemma.** *Let U be an open set in $\mathbb{C}^n$ and $V \subset U$ with $d(U,V) = r > 0$. For a differential operator*

$$P = \sum_{|J| \leq k} a_J \partial^J \in L(\mathcal{O}^c(U), \mathcal{O}^c(V))$$

of order k we have:

$$\|P\| \leq C \frac{k!}{r^k}$$

where $C = \sup_{|J| \leq a_k} |a_J|$.

Proof. If $z \in V$ and $f \in \mathcal{O}^c(U)$, then one has

$$f(z) = \frac{1}{(2\pi i)^d} \int_{\gamma_z} \frac{f(\xi)}{\prod_{i=1}^d (\xi_i - z_i)} d\xi_1 \wedge \cdots \wedge d\xi_d,$$

where γ_z denotes the cycle defined by $|\xi_i - z_i| = r$. We write it symbolically as

$$f(z) = \frac{1}{(2\pi i)^d} \int_{\gamma_z} \frac{f(\xi)}{(\xi - z)} d\xi,$$

Differentiation under the integral sign leads to

$$\partial^I f(z) = \frac{I!}{(2\pi i)^d} \int_{\gamma_z} \frac{f(\xi)}{(\xi - z)^{1+I}} d\xi.$$

We parametrise γ_z by:

$$\theta \mapsto \xi(\theta) := z + re^{\langle 2\pi i, \theta \rangle}$$

and thus

$$d\xi = (2\pi i r)^d e^{\langle 2\pi i, \theta \rangle} d\theta,$$

so that

$$\partial^I f(z) = \frac{I!}{r^{|I|}} \int_0^1 \frac{f(\xi(\theta))}{e^{2\pi i \theta}} d\theta,$$

so finally

$$|\partial^I f| \leq \frac{I!}{r^{|I|}} |f|.$$

From this the lemma follows. □

For the special case of Hamiltonian derivations, we deduce that *the norm of the map*

$$\mathcal{O}^c(U) \longrightarrow L(\mathcal{O}^c(U), \mathcal{O}^c(V)), \ h \mapsto \{h, -\}$$

is bounded by d/r^2.

5.2. **Local equivalence lemma.** *Let* $V \subset U \subset \mathbb{C}^d$ *be such that* $d(V,U) = r > 0$. *Then the restriction mapping*

$$\rho^{hc} : \mathcal{O}^h(U) \longrightarrow \mathcal{O}^c(V)$$

has norm smaller than $\pi^{-d/2} r^{-d}$ Furthermore, the canonical map

$$\rho^{ch} : \mathcal{O}^c(U) \longrightarrow \mathcal{O}^h(U)$$

has its norm bounded by $\mathrm{Vol}(U)$.

Proof. Let $f \in \mathcal{O}^h(U)$ and $w \in V$. The Taylor expansion of f at a point w

$$f(z) = \sum_{J \in \mathbb{N}^d} a_J (z - w)^J, \ a_J \in \mathbb{C}.$$

The polydisc D_w centred at w with radius r is contained in U. We have

$$\int_{D_w} |f|^2 = \sum_{J \in \mathbb{N}^d} C(J) |a_J|^2 r^{2|J| + 2n}, \quad C(I) = \prod_{k=1}^{d} \frac{\pi}{j_k + 1}.$$

So we obtain

$$C(0)|a_0|^2 r^{2n} \leq \int_{D_w} |f|^2 \leq \int_U |f|^2 = |f|^2.$$

This shows that

$$|f(w)| = |a_0| \leq \frac{c}{r^d}|f|, \quad c := \sqrt{\frac{1}{C(0)}} = \pi^{-d/2}.$$

As the point $w \in V$ was general the result follows. $\square$

5.3. **Composition lemma.** Consider the family of polydiscs $U_t = \{z \in \mathbb{C}^n \mid |z_i| < t\}$ and the Banach spaces $\mathcal{O}^c(U_t)$. Often linear operators require an arbitrary small shrinking of domain and we end up with linear maps

$$u(t,s) \in L(\mathcal{O}^c(U_t), \mathcal{O}^c(U_s))$$

defined only for $s < t$. If such a collection $u = (u(t,s), t > s)$ of linear operators is compatible with the restrictions $\mathcal{O}^c(U_{t'}) \longrightarrow \mathcal{O}^c(U_t)$ $(t' \geq t)$ and $\mathcal{O}^c(U_s) \longrightarrow \mathcal{O}^c(U_{s'})$ $(s \geq s')$, we say that u is a *horizontal section*. The *composition lemma* says:

The composition of horizontal sections u and v defined by

$$uv(t,s) = u(t, \frac{t+s}{2})v(\frac{t+s}{2}, s)$$

is again horizontal and satisfies

$$\|uv(t,s)\| \leq \|u(t, \frac{t+s}{2})\| \, \|v(\frac{t+s}{2}, s)\|$$

The proof is immediate. Often one has estimates for $u(t,s)$ and $v(t,s)$ involving only the difference $(t-s)$. The fact that we take the midpoint then leads to powers of 2 in the estimate of the composition. However, to define the composition one could take any point between t and s. For example, the the m-fold composition of horizontal sections $u_1, u_2, \ldots, u_m$ could be defined by repeated composition of two factors, but is more conveniently defined as

$$u_1 u_2 \cdots u_m(t,s) := u_1\left(t, \frac{(m-1)t+s}{m}\right) \cdots u_m\left(\frac{t+(m-1)s}{m}, s\right),$$

which often leads to powers of m in estimates, as can be observed for example in the next lemma.

5.4. **Borel lemma.** If for $t > s$ we have linear operators $u(t,s) \in L(\mathcal{O}^c(U_t), \mathcal{O}^c(U_s))$ that define a horizontal section $u = (u(t,s), t > s)$ in the sense described above. The sections for which the quantity

$$\|u\| = \sup_{s<q<p<t} \{(p-q)|u(x)|_q/(e|x|_p)\}$$

is well-defined and finite are called 1-*local*. They form a Banach space. The normalising constant $e \approx 2.718$ is purely conventional.

By composition we can form horizontal sections $u^2, u^3, u^4, \dots$. So one may try to develop a functional calculus for 1-local operators. For this to work one has to invoke a Borel transform.

If $f = \sum_{n\geq 0} a_n z^n \in \mathbb{C}\{z\}$ is an analytic series, we define its *Borel transform* as

$$\mathcal{B}f := \sum_{n\geq 0} \frac{a_n}{n!} z^n,$$

and *absolute value* as

$$|f| = \sum_{n\geq 0} |a_n| z^n.$$

Let $f = \sum_{n\geq 0} a_n z^n \in \mathbb{C}\{z\}$ a power series with R as radius of convergence and $u = (u(t,s), t > s)$ a 1-local horizontal section.

If $\|u\| < R(t-s)$, then the series $\mathcal{B}f(u)(t,s)$ converges in the operator norm and one has the estimate

$$\|\mathcal{B}f(u(t,s))\| \leq |f| \left(\frac{\|u\|}{t-s} \right).$$

Proof. From the composition lemma we find

$$\|u^n(t,s)\| \leq \|u\|^n e^{-n} n^n (t-s)^n \leq (\|u\|^n n! (t-s)^n,$$

where we used the standard inequality $n^n \leq e^n n!$ and therefore

$$\| \sum_{n\geq 0} \frac{a_n}{n!} u^n(t,s) \| \leq \sum_{n\geq 0} |a_n| \left(\frac{\|u\|}{(t-s)} \right)^n.$$

This proves the lemma. □

5.5. **Product lemma.** Let E_n be an increasing Banach scale and $(t := t_0, t_1, t_2, \dots)$ a decreasing sequence converging to $s > 0$. For any sequence (u_n) of 1-local operators $u_n \in L(E_n, E_{n+1})$, such that

i) $\|u_n\| < t_n - t_{n+1}$,
ii) $\sigma := \sum_{n\geq 0} ||u_n||/(t_n - t_{n+1}) < +\infty$,

the sequence

$$g_0 = e^{u_0}, \quad g_1 = e^{u_1} e^{u_0},$$

$$g_n := e^{u_n} e^{u_{n-1}} \cdots e^{u_0}$$

converges to an element g in the Banach space $L(E_{t_0}, E_s)$ (with operator norm). Furthermore, we have the estimate:

$$|g| < \frac{1}{1 - \sigma/(t-s)}.$$

Proof. We have seen that e^{u_i} exists as a section and thus defines elements of the Banach space $L(E_{t_i}, E_{t_{i+1}})$ as long as $\|u_i\| \leq t_i - t_{i+1}$, which holds by the first assumption. As a consequence the compositions

$$e^{u_0}, \;\; e^{u_1}e^{u_0}, \;\; \dots, e^{u_n}\dots e^{u_1}e^{u_0}, \dots$$

are well defined. Furthermore, the Borel estimate gives

$$|e^{u_i}| \leq \frac{1}{1-\nu_i}, \;\; \nu_i := \|u_i\|/(t_i - t_{i+1}).$$

As

$$\frac{1}{1-x} \times \frac{1}{1-y} < \frac{1}{1-(x+y)}$$

for $x, y \in]0,1[$, we get for the composition $e^{u_{i+1}}e^{u_i}$

$$|e^{u_{i+1}}e^{u_i}| \leq \frac{1}{1-(\nu_i+\nu_{i+1})}.$$

By a straighforward induction (and the fact that restrictions have norm ≤ 1), we obtain the estimate

$$|g_n| \leq \frac{1}{1-(\sum_{i=0}^{n}\nu_i)}.$$

Therefore

$$|g_{n+1} - g_n| \leq \frac{|e^{u_{n+1}} - \mathrm{Id}\,|}{1-(\sum_{i=0}^{n}\nu_i)}$$

Using again the Borel estimate

$$|e^{u_{n+1}} - 1| \leq \frac{\nu_{n+1}}{1-\nu_{n+1}}$$

we get

$$|g_{n+1} - g_n| \leq \frac{\nu_{n+1}}{1-(\sum_{i=0}^{n+1}\nu_i)}.$$

From this it follows that the sequence g_n converges in the Banach space $L(E_t, E_s)$ with operator norm. □

5.6. **Arnold-Moser lemma.** Let, as before, U_t denote the open polydisc of radius t and let

$$\rho(t,s) : \mathcal{O}^h(U_t) \longrightarrow \mathcal{O}^h(U_s)$$

be the restriction mappings.
The following simple result appears in [1] and [26] and is of great use:
Let $f \in \mathcal{O}^h(U_t)$ be such that its Taylor series expansions starts at order N:

$$f(z) := \sum_{|I| \geq N} a_I z^I.$$

then:

$$|\rho(t,s)f| \leq \left(\frac{s}{t}\right)^{d+N} |f|.$$

Proof. The monomials z^I form an orthogonal basis of $\mathcal{O}^h(U_s)$ with norms

$$|z^I| = C(I)^{1/2}s^{d+|I|},\ C(I) := \frac{\pi^d}{\prod_{k=1}^d (1+i_k)}.$$

By the Pythagorean theorem, for $f \in \mathcal{O}^h(U_t)$, we have:

$$\begin{aligned} |\rho(t,s)f|^2 &= \sum_{|I|\geq N} |a_I|^2 C(I) s^{2d+2|I|} \\ &= \sum_{|I|\geq N} |a_I|^2 C(I) \frac{s^{2d+2|I|}}{t^{2d+2|I|}} t^{2d+2|I|} \\ &\leq \frac{s^{2d+2N}}{t^{2d+2N}} |f|^2. \end{aligned}$$

□

5.7. **Approximation lemma.** A simple consequence of the Arnold-Moser lemma is the the following: *Let a sequence (f_N) converge to f in $\mathcal{O}^h(U_t)$. Consider the polynomials p_N obtained by truncating f_N at degree N as analytic functions in U_t. Then the sequence of polynomial (p_N) converges on any smaller polydisc to the the same limit as the restriction of f.*

Proof.

$$\begin{aligned} \|\rho(t,s)(p_N - f)\| &\leq \|\rho(t,s)(p_N - f_N)\| + \|\rho(t,s)(f_N - f)\| \\ &\leq \left(\frac{s}{t}\right)^{n+N} \|f_N\| + \|\rho(t,s)(f_N - f)\| \xrightarrow[N \longrightarrow +\infty]{} 0 \end{aligned}$$

□

5.8. **Pöschel's lemma.** For an increasing Banach scale indexed by $n \in \mathbb{N} \cup \{+\infty\}$

$$E_1 \hookrightarrow E_2 \hookrightarrow \dots E_n \longrightarrow \dots \hookrightarrow E_\infty,$$

we say that a sequence $(f_n),\ f_n \in E_n$ is *convergent*, if it maps to a converging sequence in the Banach space E_∞. A decreasing sequences of open sets

$$U_1 \supset U_2 \supset \dots \supset U_n \supset \dots \supset U_\infty := \cap_n U_n$$

gives such a Banach scale with

$$E_n := \mathcal{O}^k(U_n),$$

where $\mathcal{O}^k(U)$ is the Banach space function of complex valued C^k-function on $\overline{U}$, with bounded C^k norm which are holomorphic on the interior of U. Consider a second such decreasing sequence

$$V_1 \supset V_2 \supset \dots \supset V_n \supset \dots \supset V_\infty := \cap_n V_n$$

with $V_n \subset U_n$. The following result goes back to [30]:

Let $d(U_n, V_n) > r_n$. If a sequence (f_n), $f_n \in \mathcal{O}^c(U_n)$ converges to a limit f in $\mathcal{O}^c(U_\infty)$ faster than (r_n^k), then the sequence of restrictions $\rho_n(f_n) \in \mathcal{O}^c(V_n)$ converge to the restriction of f in $\mathcal{O}^k(V_\infty)$

Proof. By the Cauchy-Nagumo lemma, the restriction maps

$$\rho_n : \mathcal{O}^c(U_n) \longrightarrow \mathcal{O}^k(V_n)$$

satisfy the estimate

$$\|\rho_n\| \leq \frac{k!}{r_n^k}.$$

Therefore

$$\|\rho_\infty(f_n - f)\| \leq \frac{k!\|f_n - f\|}{r_n^k}.$$

Consequently the condition (for fixed k):

$$\|f_n - f\| = o(r_n^k)$$

implies the convergence of $\rho_n(f_n)$. □

References

1. V.I. Arnold, *Proof of a theorem of A. N. Kolmogorov on the preservation of conditionally periodic motions under a small perturbation of the Hamiltonian*, Uspehi Mat. Nauk **18** (1963), no. 5, 13–40, English translation: Russian Math. Surveys.
2. A.D. Bruno, *Analytic form of differential equations I*, Trans. Moscow Math. Soc. **25** (1971), 131–288.
3. R. De la Llave, *A tutorial on KAM theory*, Proceedings of Symposia in Pure Mathematics, vol. 69, Providence, RI; American Mathematical Society; 1998, 2001, pp. 175–296.
4. H. Eliasson, *Normal forms for Hamiltonian systems with Poisson commuting integrals-elliptic case*, Commentar. Math. Helv. **65** (1990), no. 1, 4–35.
5. H. Eliasson, B. Fayad, and R. Krikorian, *KAM-tori near an analytic elliptic fixed point*, Regular and Chaotic Dynamics **18** (2013), no. 6, 801–831.
6. H Eliasson, B. Fayad, and R. Krikorian, *Around the stability of KAM tori*, Duke Mathematical Journal **164** (2015), no. 9, 1733–1775.
7. C. Fefferman, *Extension by linear operators*, Annals of mathematics (2007), 779–835.
8. J. Féjoz and M. Garay, *Un théorème sur les actions de groupes de dimension infinie*, Comptes Rendus à l'Académie des Sciences **348** (2010), no. 7-8, 427–430.
9. M. Garay, *Stable moment mappings and singular Lagrangian fibrations*, Quarterly Journal of Mathematics **56** (2005), 357–366.
10. ______, *General KAM theorems*, vol. 6, Oberwolfach reports, no. 3, European Mathematical Society, 2009, pp. 2443–2446.
11. ______, *The Herman conjecture*, ArXiv: 1206.1245 (unpublished), 2012.
12. ______, *The Herman conjecture*, Oberwolfach reports, European Mathematical Society, 2012.
13. ______, *An Abstract KAM Theorem*, Moscow Mathematical Journal **14** (2014), no. 4, 745–772.

14. ______, *Degenerations of invariant Lagrangian manifolds*, Journal of Singularities **8** (2014), 50–67.
15. ______, *Arithmetic Density*, Proceedings of the Edinburgh Mathematical Society **59** (2016), no. 3, 691–700.
16. M. Garay and D. van Straten, *KAM Theory*, ArXiv: 1204.2493, 2018.
17. A. González-Enríquez, A. Haro, and R. De la Llave, *Singularity theory for non-twist KAM tori*, vol. 227, Memoirs of the AMS, no. 1067, American Mathematical Society, 2014.
18. M.R. Herman, *Some open problems in dynamical systems*, Proceedings of the International Congress of Mathematicians, Doc. Math., vol. II, 1998, pp. 797–808.
19. D. Kleinbock, *Extremal subspaces and their submanifolds*, Geom. Funct. Anal **13** (2003), no. 2, 437–466.
20. D.Y. Kleinbock and G.A. Margulis, *Flows on homogeneous spaces and Diophantine approximation on manifolds*, Ann. of Math. **148** (1998), 339–360.
21. A.N. Kolmogorov, *On the conservation of quasi-periodic motions for a small perturbation of the Hamiltonian function*, Dokl. Akad. Nauk SSSR **98** (1954), 527–530.
22. B. Malgrange, *The preparation theorem for differentiable functions*, Differential Analysis, Bombay Colloq, 1964, pp. 203–208.
23. J. Mather, *Stability of C^∞ mappings. I. The division theorem*, Annals of Math. **87** (1968), 89–104.
24. E. Miranda and S. Vũ Ngoc, *A Singular Poincaré Lemma* , Int. Math. Research Notices **58A10** (2005), no. 1, 27–45.
25. J. Moser, *On the generalization of a theorem of A. Liapounoff*, Communications on Pure and Applied Mathematics **11** (1958), no. 2, 257–271.
26. ______, *A rapidly convergent iteration method and non-linear partial differential equations I*, Ann. Scuola Norm Sup. Pisa - Classe di Scienze Sér. 3 **20** (1966), no. 2, 265–315.
27. ______, *On the construction of almost periodic solutions for ordinary differential equations (Tokyo, 1969)*, Proc. Internat. Conf. on Functional Analysis and Related Topics, Univ. of Tokyo Press, 1969, pp. 60–67.
28. A. Ottolenghi, *On convergence of normal forms for complex frequencies*, Journal of mathematical physics **34** (1993), no. 11, 5205–5220.
29. R. Pérez-Marco, *Convergence or generic divergence of the Birkhoff normal form*, Ann. of Math. **2** (2003), 557–574.
30. J. Pöschel, *Integrability of Hamiltonian systems on Cantor sets*, Comm. Pure Appl . Math. **35** (1982), 653–695.
31. ______, *A lecture on the classical KAM theorem*, Smooth Ergodic Theory and Its Applications (A.B. Katok, R. de la Llave, Ya.B. Pesin, and H.Weiss, eds.), Proceedings of Symposia in Pure Mathematics, vol. 69, Amer. Math. Soc., 2001, pp. 707–732.
32. H. Rüssmann, *Über das Verhalten analytischer Hamiltonscher Differentialgleichungen in der Nähe einer Gleichgewichtslösung*, Math. Annalen **154** (1964), 285–306.
33. ______, *Nondegeneracy in the perturbation theory of integrable dynamical systems*, Number theory and dynamical systems (York, 1987), London Math. Soc., Cambridge University Press, 1989, pp. 5–18.
34. M.B. Sevryuk, *The classical KAM theory at the dawn of the twenty-first century*, Moscow Math. Journal **3** (2003), no. 3, 1113–1144.
35. C.L. Siegel, *On the integrals of canonical systems*, Annals of Mathematics **42** (1941), no. 3, 806–822.

36. L. Stolovitch, *Singular complete integrability*, Inst. Hautes Études Sci. Publ. Math **91** (2000), 133–210.
37. J. Vey, *Sur certains sytèmes dynamiques séparables*, American Journal of Math **100** (1978), 591–614.
38. H. Whitney, *Analytic extensions of differentiable functions defined in closed sets*, Trans. Amer. math. Soc. **36** (1934), 63–89.
39. N.T. Zung, *Convergence versus integrability in Birkhoff normal form*, Ann. of Math. **161** (2005), no. 1, 141–156.