Hamiltonian perturbation theory
for ultra-differentiable functions

Abed Bounemoura

Jacques Féjoz

Author address:

PSL RESEARCH UNIVERSITY, CNRS, CEREMADE AND IMCCE, PARIS, FRANCE
E-mail address: abedbou@gmail.com, jacques.fejoz@dauphine.fr

Contents

Abstract

Some scales of spaces of ultra-differentiable functions are introduced, having good stability properties with respect to infinitely many derivatives and compositions. They are well-suited for solving non-linear functional equations by means of hard implicit function theorems. They comprise Gevrey functions and thus, as a limiting case, analytic functions. Using majorizing series, we manage to characterize them in terms of a real sequence M bounding the growth of derivatives.

In this functional setting, we prove two fundamental results of Hamiltonian perturbation theory: the invariant torus theorem, where the invariant torus remains ultra-differentiable under the assumption that its frequency satisfies some arithmetic condition which we call BR_M, and which generalizes the Bruno-Rüssmann condition; and Nekhoroshev's theorem, where the stability time depends on the ultra-differentiable class of the perturbation, through the same sequence M. Our proof uses periodic averaging, while a substitute for the analyticity width allows us to bypass analytic smoothing.

We also prove converse statements on the destruction of invariant tori and on the existence of diffusing orbits with ultra-differentiable perturbations, by respectively mimicking a construction of Bessi (in the analytic category) and Marco-Sauzin (in the Gevrey non-analytic category). When the perturbation space satisfies some additional condition (we then call it *matching*), we manage to narrow the gap between stability hypotheses (e.g. the BR_M condition) and instability hypotheses, thus circumscribing the stability threshold.

The formulas relating the growth M of derivatives of the perturbation on the one hand, and the arithmetics of robust frequencies or the stability time on the other hand, bring light to the competition between stability properties of nearly integrable systems and the distance to integrability. Due to our method of proof using width of regularity as a regularizing parameter, these formulas are closer to optimal as the the regularity tends to analyticity.

Introduction

1. High regularity in perturbation theory

The purpose of this article is two-fold. First, we introduce and study the general properties of a class of ultra-differentiable functions well-suited for solving non-linear functional equations with iterative methods. Second, within this framework, we generalize the fundamental results of Hamiltonian perturbation theory such as the KAM theorem on the preservation of invariant tori or the Nekhoroshev theorem on the long-time stability of solutions; we also prove "converse" statements such as the destruction of tori or the construction of unstable orbits (the so-called Arnold diffusion).

First recall that real-analytic functions in m variables are characterized by a growth of their derivatives of order $s^{-|k|}|k|!$ for $k \in \mathbb{N}^m$ and some analyticity width $s > 0$; in the periodic case, this is equivalent to a decay of Fourier coefficients of order $e^{-s|j|}$ for $j \in \mathbb{Z}^m$.

In his foundational paper [**Kol54**], Kolmogorov proved that an invariant torus in a real-analytic non-degenerate integrable Hamiltonian system can be preserved, as a real-analytic embedded torus, under an arbitrary small real-analytic perturbation, provided the frequency vector satisfies a Diophantine condition. This theorem of Kolmogorov has generated a tremendous amount of work, and new proofs of more general statements appeared. Among them, still in the analytic case, by using very accurate approximation results of analytic functions by polynomials, Rüssmann ([**Rüs01**]) was able to improve the arithmetic condition, and get what is now called the Bruno-Rüssmann condition (BR-condition for short). This condition, which was introduced earlier (in a different form) by Bruno ([**Bru71**], [**Bru72**]) in the Siegel linearization problem (a different problem involving small divisors), is known to be optimal in the Siegel problem in complex dimension one (this is a celebrated result of Yoccoz [**Yoc88**], [**Yoc95**]). Whether BR is optimal in higher dimension or in the Hamiltonian problem is open. A different proof of the Kolmogorov theorem with the BR-condition was then given in [**BF13**], [**BF14**]; instead of approximating the real-analytic perturbation by a polynomial, the idea consists of approximating the frequency vector by suitable periodic vectors. Let us point out that all these proofs crucially use a complex extension of the domain and of the Hamiltonian in order to use methods of complex analysis. Also, even though it is not known if the BR-condition is optimal, a construction of Bessi [**Bes00**], inspired by the theory of Arnold diffusion, shows that BR cannot be improved much, if at all.

A second fundamental result is due to Nekhoroshev, who showed in [**Nek77**], [**Nek79**] that all solutions of a perturbed "generic" real-analytic integrable Hamiltonian system are stable for an exponentially long interval of time. This generic class of integrable Hamiltonians, which are known as steep, includes as a particular case the quasi-convex functions for which the proof of Nekhoroshev theorem greatly simplifies. As a consequence, better quantitative results were obtained in [**Loc92**], [**LN92**], [**LNN94**] and in [**Pös93**] for the quasi-convex case. Together with the improvements obtained in [**BM11**] and in [**ZZ17**] and the examples of analytic Arnold diffusion in [**Bes96**], [**Bes97**] and [**Zha11**], all these results determine precisely the time-scale of stability. The quantitative result of the quasi-convex case was recently extended to the general steep case in [**GCB16**]; however, no examples of Arnold diffusion are known for steep non-convex integrable Hamiltonians. Let us also remark that even though linear integrable Hamiltonians are non-steep, they are stable for a long interval of time provided the (constant) frequency is non-resonant; this is a well-known result assuming the frequency to be Diophantine, but in the general case this was

observed in [**Bou12**] along with the fact that the stability estimates thus obtained are optimal as examples of Arnold diffusion with controlled speed are quite easily constructed in this context.

Next, given a real parameter $\alpha \geq 1$, allowing a growth of the derivatives of order $s^{-|k|}|k|!^\alpha$ for $k \in \mathbb{N}^m$ or, equivalently, a decay of Fourier coefficients of order $e^{-(s|j|)^{1/\alpha}}$ for $j \in \mathbb{Z}^m$ in the periodic case, one is led to consider α-Gevrey functions of m variables, which thus reduce to real-analytic functions when $\alpha = 1$. The "width" $s > 0$ extends the significance of the analyticity width of analytic functions. Since the introduction by Gevrey of the class of functions now bearing his name ([**Gev18**]), there has been a huge amount of works on Gevrey functions, mainly for PDEs, but also more recently in other fields, including dynamical systems.

The theorem of Kolmogorov, under the classical Diophantine condition, was extended to this Gevrey context by Popov [**Pop04**]. His proof, which is based on approximation of Gevrey functions by real-analytic ones, did not allow him to reach a weaker arithmetic condition even though one would expect the statement to be true under a condition which generalizes the BR-condition. Recently, we introduced in [**BF17**] such a condition that we called the BR$_\alpha$-condition and under which an invariant torus is preserved as an α-Gevrey torus for a non-degenerate close to integrable α-Gevrey system. The proof in [**BF17**] was direct, in the sense that in the real-analytic case it does not use any complex extension, and in the Gevrey case it does not use analytic approximation. Actually, one of the main difficulties was to obtain functional estimates (and in particular product, derivatives and composition estimates) which were not known and were clearly required if one wants to implement an iterative scheme (typical of KAM) within the α-Gevrey category. Indeed, of particular importance was to obtain a norm estimate for the composition of a function by a diffeomorphism in which the loss of "width" is arbitrarily small provided the diffeomorphism is arbitrarily close to the identity; such an estimate is needed if one insists on having some "width" parameter left after infinitely many compositions. These crucial estimates were obtained in [**BF17**], and with further work using the method introduced in [**BF13**], [**BF14**] the result was achieved. Also we observed that Bessi's example could be modified to give a necessary condition close to the sufficient BR$_\alpha$-condition; in particular, invariant torus with a frequency which satisfy the BR-condition can be destroyed by an arbitrarily small perturbation which is α-Gevrey for any $\alpha > 1$, and this is in sharp contrast with some other small divisors problem (as for instance the formal Gevrey Siegel problem considered in [**CM00**] and [**Car03**]).

As for the Nekhoroshev theorem, the extension to α-Gevrey quasi-convex Hamiltonians was achieved in [**MS02**]. Here precise composition estimates were not needed (even though the authors were aware that their composition result was not accurate). Indeed, it is known since the work of Lochak that for quasi-convex Hamiltonians, one only needs a normal form at a periodic frequency and thus, from an analytical point of view, one is dealing with the problem of eliminating the time-dependence in a slow-fast time-periodic system. This problem always has a formal solution which of course may be divergent: in [**RS96**] it was proved that for an analytic system, the associated formal series has a Gevrey character, and this was extended (in a Hamiltonian setting) for an arbitrary α-Gevrey system. Once one knows the existence of a formal solution with such a Gevrey Character, techniques of re-summation allow to find a Gevrey (convergent) normal form with an exponentially small remainder, and this is all what is needed to prove the Nekhoroshev theorem for quasi-convex Hamiltonians. Examples of Arnold diffusion were also obtained in [**MS02**] but in the Gevrey non-analytic case $\alpha > 1$, as the method uses the existence of bump functions. However, even though the improvement of [**BM11**] does extend to the more general Gevrey setting, for the moment this is not the case of [**ZZ17**] and thus the optimal time-scale of stability in the Gevrey case is not yet completely determined. Nekhoroshev estimates have been obtained also in the general steep case for Gevrey Hamiltonians (see [**Bou11**]), yet the quantitative result is very poor compared to the analytic case studied in [**GCB16**]. Finally, for Gevrey linear integrable Hamiltonians, precise and essentially optimal stability and instability results are proved in [**Bou13a**].

More generally, for a sequence of positive real numbers $M = (M_l)_{l \in \mathbb{N}}$, we can consider the class of smooth functions having a growth of derivatives of order $s^{-|k|} M_{|k|}$ for $k \in \mathbb{N}^m$ and some $s > 0$ which we still call a "width". This defines a class of ultra-differentiable functions associated

to M, or simply M-ultra-differentiable functions. The real-analytic and α-Gevrey regularity are nothing but the particular cases where respectively $M_l = l!$ and $M_l = l!^\alpha$. In terms of Fourier coefficients for periodic functions, this amounts to a decay of order $e^{-\Omega(s|j|)}$ for $j \in \mathbb{Z}^m$, for some suitable positive increasing function Ω; again the real-analytic and α-Gevrey regularity correspond to respectively $\Omega(y) = y$ and $\Omega(y) = y^{1/\alpha}$. It is one of our goal here to introduce a class of ultra-differentiable functions in which non-linear problems can be solved by an iterative method within this class. To do so, we will restrict the allowed sequence M by requiring two conditions to be satisfied (see § 2 for the precise definition of M-ultra-differentiable functions and these two conditions). The first of these conditions (which we will call (H1)) is a kind of regularity condition and it is classical and very often satisfied. In particular it implies stability by product and by composition; for the composition a loss of "width" occurs so in order to be able to perform infinitely many compositions, we will have to show that the loss can be arbitrarily small when the composition involves a diffeomorphism arbitrarily close to the identity. The second condition (which we will call (H2)) is, however, more restrictive as it requires the sequence M not to grow too fast at infinity; it implies not only stability by derivation but moreover that the loss of width can be taken arbitrarily small, and thus one may say that this condition amounts to stability under 'infinitely many derivations". This growth restriction is quantified by a function which we will call C, and in good cases it will be related to the function Ω which controls the decay of Fourier coefficients in the periodic case (see § 3 for more information on these functions C and Ω). The functional estimates we obtained in [**BF17**] in the Gevrey setting will be generalized to this M-ultra-differentiable setting where M satisfies these two conditions (H1) and (H2).

As a first application, we will give the proof of KAM type theorems within this class of M-ultra-differentiable functions (see § 5 for precise statements). We will introduce an arithmetic condition, the BR_M-condition, and under which an invariant torus is preserved as an M-ultra-differentiable torus for a non-degenerate close to integrable M-ultra-differentiable Hamiltonian system (see § 4 for the definition of this arithmetic condition). This BR_M-condition reduces to the BR_α-condition when $M_l = l!^\alpha$ so this includes the main result of [**BF17**] but it encompasses many more cases. Let us point out that this BR_M-condition does not depend directly on M but rather on the associated function C that was mentioned above. Again, Bessi's example could be modified to give a necessary condition for the preservation of an invariant torus, yet this necessary condition now involves the function Ω and not C and thus only in the good cases (where these two functions are related; this will be called *matching* in the sequel) this necessary condition shows that the sufficient condition cannot be improved much, if at all.

As a second application, we will extend Nekhoroshev theory to this class of M-ultra-differentiable functions (see § 6 for precise statements). As we already explained, in the analytic case we have a good knowledge, the main issue being the Arnold diffusion for steep non-convex systems. In the more general Gevrey case, the time-scale of stability is not yet completely determined in the quasi-convex case but also only poor quantitative statements are known in the general steep case. In the ultra-differentiable setting, basically we will be able to extend the knowledge we have for the general Gevrey case but not the more precise knowledge of the analytic case. First, for linear integrable Hamiltonians, we will give accurate and essentially optimal stability and instability results, which completely extend the analytic and Gevrey cases; exactly like for the KAM type results, the stability (that is, positive) result involves the C function whereas the diffusion (that is, negative) result involves the Ω function and so one has to consider the result to be accurate only for matching sequences. Then, for quasi-convex integrable Hamiltonians, our stability result again will be a complete generalization of the analytic and Gevrey cases, however, exactly like for the Gevrey case, the improvement of [**ZZ17**] cannot be reached and the construction of Arnold diffusion, following [**MS02**], requires bump functions and thus a further assumption of non-quasi-analyticity. Finally, in the steep case, a result generalizing [**Bou11**] will be obtained but the quantitative result will be again quite far from the more accurate analytic case [**GCB16**].

2. Ultra-differentiable functions

Recall that $n \geq 1$ is an integer, $\mathbb{T}^n = \mathbb{R}^n / \mathbb{Z}^n$ and let $D \subseteq \mathbb{R}^n$ be the open ball of radius one centered at the origin. For a small parameter $\varepsilon \geq 0$, we consider a Hamiltonian function $H : \mathbb{T}^n \times D \to \mathbb{R}$ of the form

$$(*) \qquad \begin{cases} H(\theta, I) = h(I) + \varepsilon f(\theta, I), \\ \nabla h(0) := \omega_0 \in \mathbb{R}^n. \end{cases}$$

We shall assume that the Hamiltonian H is *ultra-differentiable* on $\mathbb{T}^n \times D$, in the following sense. Let us fix a sequence $M := (M_l)_{l \in \mathbb{N}}$ of positive real numbers with $M_0 = M_1 = 1$. The function H is said to be ultra-differentiable with respect to M (or M-ultra-differentiable) if H is smooth and there exists $s > 0$ such that, using multi-indices notation (see § 1),

$$(1) \qquad |H|_{M,s} := c \sup_{(\theta, I) \in \mathbb{T}^n \times D} \left(\sup_{k \in \mathbb{N}^{2n}} \frac{(|k|+1)^2 s^{|k|} |\partial^k H(\theta, I)|}{M_{|k|}} \right) < \infty, \quad c := 4\pi^2/3.$$

The space of such Hamiltonians will be denoted by $\mathcal{U}_{M,s}(\mathbb{T}^n \times D)$ or, more simply, $\mathcal{U}_s(\mathbb{T}^n \times D)$. This definition can be extended to vector-valued functions $X : \mathbb{T}^n \times D \to \mathbb{R}^p$ by setting

$$(2) \qquad |X|_{M,s} := c \sup_{(\theta, I) \in \mathbb{T}^n \times D} \left(\sup_{k \in \mathbb{N}^{2n}} \frac{(|k|+1)^2 s^{|k|} |\partial^k X(\theta, I)|_1}{M_{|k|}} \right) < \infty$$

where $|\,.\,|_1$ is the l_1-norm of vectors in $\mathbb{R}^p$, or the sum of the absolute values of the components, thus defining the space $\mathcal{U}_s(\mathbb{T}^n \times D, \mathbb{R}^p)$. As a rule, we will use the l_1-norm for vectors, so for simplicity we shall write $|\,.\,|_1 = |\,.\,|$.

Condition (1) may be seen as some modification of the Taylor series of H having radius of convergence at least s. Thus, the parameter s extends the concept of width of analyticity of an analytic function (see example 1), and will be called the (ultra-differentiable) *width*. To emphasize its role, we shall also say that H is (M, s)-ultra-differentiable if (1) holds. These classes of ultra-differentiable functions we are considering here are usually called Denjoy-Carleman classes (of Roumieu type) in the literature; we refer to [**Thi08**] for a nice survey.

General properties of these ultra-differentiable norms (under certain assumptions we will introduce below) are described in § 1. In particular we explain there the (inessential) role of the factor $(|k|+1)^2$ and the normalizing constant $c > 0$ in (1).

Let us now give the main example of ultra-differentiable class of functions.

EXAMPLE 1. Given $\alpha \geq 1$, let us define the sequence

$$(3) \qquad M_\alpha := (l!^\alpha)_{l \in \mathbb{N}}.$$

Observe that a function is M_α-ultra-differentiable if and only if it is α-Gevrey, and thus it is M_1-ultra-differentiable if and only if it is real-analytic, in which case the parameter $s > 0$ is the width of analyticity. (The case when $0 \leq \alpha < 1$ corresponds to a class of entire functions and will not be considered in this work.) More examples of sequences will be given in § 4.

Without further information on the sequence M, one cannot expect much structure on the space of M-ultra-differentiable functions. For our purposes here, we will have to make two assumptions. But first associated with the sequence M we define three other sequences $\mu := (\mu_l)_{l \in \mathbb{N}}$, $N := (N_l)_{l \in \mathbb{N}}$ and $\nu := (\nu_l)_{l \in \mathbb{N}}$ by

$$(4) \qquad \mu_l := \frac{M_{l+1}}{M_l}, \quad N_l := \frac{M_l}{l!}, \quad \nu_l := \frac{N_{l+1}}{N_l} = \frac{\mu_l}{l+1}.$$

As $M_0 = M_1 = 1$, we have $N_0 = N_1 = 1$ and $\mu_0 = \nu_0 = 1$. Clearly, the knowledge of one of these sequences determines the other three. The two assumptions we will require are the following:

(H1)   The sequence N is log-convex, i.e., the sequence ν is non-decreasing.

(H2)   The sequence μ is sub-exponential, i.e., $\displaystyle \lim_{l \to +\infty} l^{-1} \ln(\mu_l) = 0$.

Let us briefly describe the role of these two assumptions.

The log-convexity condition (H1) is a very classical assumption in the literature, in particular it implies that the space of M-ultra-differentiable function is closed under product and composition; as a matter of fact, for this to be true slightly weaker conditions are required (see Lemma 12 and Lemma 15 in respectively § 4 and § 5). Clearly, if ν is increasing then so is μ thus the log-convexity of N implies the log-convexity of M (but the reverse implication is not true in general). Let us also point out that (H1) and the normalization $N_0 = M_0 = 1$ imply

$$N_l \geq 1, \quad M_l \geq l!, \quad l \in \mathbb{N}$$

with the limiting case where $N_l = 1$, $M_l = l!$, corresponding to the space of analytic functions.

Up to our knowledge, the sub-exponential condition (H2) hasn't been used so far in the literature. This condition implies that our space of functions is closed under derivation, which is easily seen to be equivalent to the condition

$$\sup_{l \in \mathbb{N}} \mu_l^{\frac{1}{l}} = \sup_{l \in \mathbb{N}} \left(\frac{M_{l+1}}{M_l} \right)^{\frac{1}{l}} < +\infty,$$

but it is definitely stronger: instead of just requiring that the sequence $\left(\mu_l^{1/l} \right)_{l \in \mathbb{N}}$ is bounded, (H2) demands that this sequence converges to 1.

However, (H2) is not an inordinate property. Indeed, a condition commonly assumed in the literature on ultra-differentiable functions is the following, of so-called *moderate growth*:

(MG) $$\sup_{l,j \in \mathbb{N}} \left(\frac{M_{l+j}}{M_l M_j} \right)^{\frac{1}{l+j}} < +\infty.$$

As Lemma 46 shows (Appendix A), this condition is stronger than our hypothesis (H2).

Actually, one can see (H2) as the guarantee that our functional space is "stable by infinitely many derivations". Indeed, in view of our choice of norms, the condition (H2) is exactly the one that ensures we have an analogue of the Cauchy estimates for analytic functions: if f is (M, s)-ultra-differentiable, then its derivative f' is (M, s')-ultra-differentiable for any $s' < s$ and we can bound the (M, s')-norm of f' in terms of the (M, s)-norm of f and a function of $s - s'$, that we call $C(s - s')$ (C stands for Cauchy). As s' converges to s, this function blows up and its asymptotics play an important role in the sequel. It will be defined and briefly studied in the next section, along another function, which we call Ω (this function was already mentioned in the introduction) and which can actually be defined without any assumption on the sequence M.

3. Functions C and Ω

Given $0 < \sigma < 1$, we define the Cauchy function $C = C_M$ by

(5) $$C(\sigma) := \sup_{l \in \mathbb{N}} \mu_l e^{-\sigma l} = \sup_{l \in \mathbb{N}} \frac{M_{l+1}}{M_l} e^{-\sigma l}$$

where the sequence μ was defined in (4): that $C(\sigma)$ is finite for all $\sigma > 0$ follows from (and in fact is equivalent to) (H2) which guarantees that μ has a sub-exponential growth. Let us first remark that $\nu_l \geq \nu_0 = 1$ (this follows from (H1) and the normalization) and thus

(6) $$C(\sigma) = \sup_{l \in \mathbb{N}} (l+1) \nu_l e^{-\sigma l} \geq \sup_{l \in \mathbb{N}} (l+1) e^{-\sigma l} \geq (e\sigma)^{-1}.$$

Clearly, the supremum in (5) is in fact a maximum and the function C is thus continuous on $]0, 1[$. Besides, there exists some $0 < \bar{\sigma} < 1$ which depends only on the sequence M such that $C(\sigma) = 1$ for any $\bar{\sigma} \leq \sigma < 1$ (we will assume that $\bar{\sigma}$ is minimal with this property); the interesting limit is when σ goes to zero (in the discussion above, this corresponds to $s' = s(1-\sigma)$ and the limit $s' \to s$ amounts to $\sigma \to 0$). For $0 < \sigma \leq \bar{\sigma}$, the function C is then decreasing, and so the restriction

$$C :]0, \bar{\sigma}] \to [1, +\infty[$$

is a homeomorphism, which has a continuous decreasing functional inverse

$$C^{-1} : [1, +\infty[\to]0, \bar{\sigma}].$$

The behavior of $C(\sigma)$, as σ tends to zero, translates into the behavior of $C^{-1}(y)$, as y tends to infinity.

Let us now introduce another function, this one classically attached to the sequence M, which we will call $\Omega = \Omega_M$, and which is defined by

$$(7) \qquad \Omega(y) := \ln\left(\sup_{l \in \mathbb{N}} y^l M_l^{-1}\right) = \sup_{l \in \mathbb{N}}\left(l \ln y - \ln M_l\right).$$

One can show (see [CC94] or [Thi03]) that this defines a function

$$\Omega : [0, +\infty[\to [0, +\infty[$$

which is continuous, constant equal to zero for $y \leq 1$ and strictly increasing for $y \geq 1$. Clearly, $\Omega(y)$ grows faster than $\ln\left(y^l\right)$ for any $l \in \mathbb{N}$ as y goes to infinity, but also

$$(8) \qquad \lim_{y \to +\infty} \frac{\Omega(y)}{\ln(y)} = +\infty.$$

For a periodic function $f \in \mathcal{U}_s(\mathbb{T}^n)$, letting $f = \sum_{j \in \mathbb{Z}^n} f_j e_j$ be its Fourier expansion, one easily checks that Ω controls the decay of the Fourier coefficients in the sense that

$$(9) \qquad |f_j| \leq \exp(-\Omega(s|j|)), \quad j \in \mathbb{Z}^n.$$

Here too, it is the asymptotic of the function Ω at infinity which is of interest to us, so to make things more precise, let us introduce the following definition.

Let us say that two real-valued functions f and g of one variable, defined in the neighborhood of u, $u \in \{0, +\infty\}$, are *equivalent up to scalings* (at u) if there exist positive constants a and b such that

$$f(ax) \leq g(x) \leq f(bx), \quad x \to u.$$

In the sequel we shall simply write $f(x) \asymp g(x)$ without referring to the precise value u.

EXAMPLE 2. Let us come back to the Example 1 where

$$M_\alpha = (l!^\alpha)_{l \in \mathbb{N}}, \quad \mu_\alpha = ((l+1)^\alpha)_{l \in \mathbb{N}}.$$

Clearly, (H1) and (H2) are satisfied, and simple computations lead to

$$C(\sigma) \asymp \sigma^{-\alpha}, \quad C^{-1}(y) \asymp y^{-1/\alpha}, \quad \Omega(y) \asymp y^{1/\alpha}.$$

In the above example, one remarks that the inverse of Ω is equivalent up to scalings (at infinity) to the functional inverse of C. Let us give a name to sequences with such a property: a sequence M will be called *matching* if

$$(10) \qquad C^{-1} \asymp 1/\Omega$$

where the functions C and Ω have been defined respectively in (5) and (7). In the sequel, matching sequences will play a special role: these will be the sequences for which we will be able to compare in the sharpest way stability results (KAM, Nekhoroshev) to instability results (destruction of tori, Arnold diffusion).

4. An arithmetic condition for ultra-differentiable functions

Given $\omega_0 \in \mathbb{R}^n$, define the function

$$(11) \qquad \Psi_{\omega_0} : [1, +\infty) \to [\Psi_{\omega_0}(1), +\infty], \quad Q \mapsto \max\{|k \cdot \omega_0|^{-1} \mid k \in \mathbb{Z}^n, \, 0 < |k| \leq Q\}.$$

This function Ψ_{ω_0} measures the size of the so-called small denominators. We will impose a condition that prevents Ψ_{ω_0} from growing too fast at infinity; the precise condition will actually depend on the sequence M through C^{-1}, the inverse of the Cauchy function C we introduced in the previous section.

But first let us recall that the function Ψ_{ω_0} is non-decreasing (thus with a countable number of discontinuities) and piecewise constant. In the sequel, it will be more convenient to work with a continuous version of Ψ_{ω_0}: it is not hard to prove (see, for instance, Appendix A of [BF13])

that one can find a continuous non-decreasing function $\Psi : [1, \infty) \to [\Psi(1), +\infty)$ such that $\Psi(1) = \Psi_{\omega_0}(1)$ and

$$(12) \qquad \Psi_{\omega_0}(Q) \leq \Psi(Q) \leq \Psi_{\omega_0}(Q+1), \quad Q \geq 1.$$

Clearly, for all $k \in \mathbb{Z}^n \setminus \{0\}$, we still have

$$|k \cdot \omega_0| \geq 1/\Psi(|k|)$$

and in the condition we will introduce below, one may use Ψ instead of Ψ_{ω_0}.

Let us now define the function

$$\Delta : [1, +\infty) \to [\Psi(1), +\infty), \quad Q \mapsto Q\Psi(Q),$$

which is continuous and increasing; Δ has therefore a functional inverse

$$\Delta^{-1} : [\Psi(1), +\infty) \to [1, +\infty), \quad \Delta^{-1} \circ \Delta = \Delta \circ \Delta^{-1} = \mathrm{Id}$$

which is also continuous and increasing.

We are now ready to introduce our arithmetic condition adapted to M-ultra-differentiable functions. We say that $\omega_0 \in \mathrm{BR}_M$ if, for any $c > 0$,

$$(\mathrm{BR_M}) \qquad \int_{\Delta(1/c)}^{+\infty} C^{-1}(c\Delta^{-1}(x)) \frac{dx}{x} < +\infty,$$

where C^{-1} is the functional inverse of the Cauchy function C defined in (5). This definition is somewhat involved, which calls for a few comments and examples.

First observe that the condition $(\mathrm{BR_M})$ is unchanged if one replaces C^{-1} by a function which is equivalent to it up to scalings. In particular, instead of using the function Ψ, we may as well use Ψ_{ω_0} (the role of Δ^{-1} will then be played by a generalized inverse of $Q \mapsto Q\Psi_{\omega_0}(Q)$) as one can easily check from (12) (see Appendix A of [**BF13**] for more details). The set $\mathrm{BR_M}$ may be empty (see Example 6 below) or not. Before giving concrete examples, recall that the set D_τ of τ-Diophantine vectors ($\tau \geq n - 1$) is the set of vectors for which there exists $\gamma > 0$ such that $\Psi_\omega(Q) \leq Q^\tau/\gamma$ for all $Q \geq 1$. It is well-known and easy to prove that D_τ is non-empty, and has full measure if $\tau > n - 1$.

EXAMPLE 3. Let us come back to the Examples 1 and 2, where

$$M_\alpha = (l!^\alpha)_{l \in \mathbb{N}}, \quad C^{-1}(y) \asymp y^{-1/\alpha}.$$

As we already pointed out, M_α is matching. In this case, condition $(\mathrm{BR_M})$ is satisfied if and only if

$$(\mathrm{A_\alpha}) \qquad \int_{\Delta(1)}^{+\infty} \frac{1}{x(\Delta^{-1}(x))^{1/\alpha}} dx < \infty$$

which is a condition we introduced in [**BF17**], and where it is shown to be equivalent to the α-Bruno-Rüssmann condition

$$(\mathrm{BR_\alpha}) \qquad \int_1^{+\infty} \frac{\ln(\Psi_{\omega_0}(Q))}{Q^{1+1/\alpha}} dQ < \infty.$$

Let us denote by $\mathrm{BR_\alpha}$ the set of vectors satisfying this latter condition. It is not hard to check that $\mathrm{BR_\alpha}$ is strictly included in $\mathrm{BR_{\alpha'}}$ if $\alpha' < \alpha$ and

$$\cap_{\alpha \geq 1} \mathrm{BR_\alpha} \setminus \cup_{\tau \geq n-1} D_\tau$$

is non-empty.

EXAMPLE 4. Given $\alpha \geq 1$ and $\beta \geq 0$, let us now consider more generally the sequence $M_{\alpha,\beta}$ defined by

$$(M_{\alpha,\beta})_0 = 0, \quad (M_{\alpha,\beta})_l = l!^\alpha \prod_{i=0}^l (\ln(e - 1 + i))^\beta, \quad l \geq 1.$$

Clearly, $M_{\alpha,0} = M_\alpha$ and (H1) and (H2) are satisfied. We have

$$\mu_j = (j+1)^\alpha (\ln(e+j))^\beta$$

and elementary computations show that

$$C(\sigma) \asymp \sigma^{-\alpha}(\ln(\sigma^{-1}))^{\beta}, \quad C^{-1}(y) \asymp y^{-1/\alpha}(\ln(y))^{\beta/\alpha}$$

but also

$$\Omega(y) \asymp y^{1/\alpha}(\ln(y))^{-\beta/\alpha}$$

and so $M_{\alpha,\beta}$ is matching. In this case, condition $(\mathrm{BR_M})$ is satisfied if and only if

$$(\mathrm{BR}_{\alpha,\beta}) \qquad \int_{\Delta(1)}^{+\infty} \frac{(\ln(\Delta^{-1}(x)))^{\beta/\alpha}}{x(\Delta^{-1}(x))^{1/\alpha}} dx < \infty.$$

By extension, let us denote by $\mathrm{BR}_{\alpha,\beta}$ the set of vectors satisfying this latter condition, then $\mathrm{BR}_{\alpha,\beta}$ is strictly included in $\mathrm{BR}_{\alpha',\beta'}$ if $\alpha' < \alpha$ or $\alpha' = \alpha$ and $\beta' < \beta$, and

$$\cap_{\alpha \geq 1,\, \beta \geq 0} \mathrm{BR}_{\alpha,\beta} = \cap_{\alpha \geq 1} \mathrm{BR}_\alpha \setminus \cup_{\tau \geq n-1} D_\tau.$$

EXAMPLE 5. Let us now give an example in which the growth of the sequence M is wilder. Equivalently, it is sufficient to prescribe the sequence μ, and let us choose

$$\mu_l = \exp((\ln(l))^2), \quad M_l \sim \exp(l(\ln(l))^2).$$

As before, one easily check that (H1) and (H2) are satisfied. One can then compute

$$C(\sigma) \asymp \exp\left(\left(\ln\sigma^{-1}\right)^2 + \ln\sigma^{-1}\ln\ln\sigma^{-1}\right)$$

and

$$C^{-1}(y) \asymp \exp\left(-(\ln y - \ln\ln y)^{1/2}\right).$$

Here, one can check that the sequence is not matching. But observing for instance that given any $d > 1$, for y large one has

$$C^{-1}(y) \leq (\ln y)^{-d}$$

then one easily sees that the set of vectors satisfying condition $(\mathrm{BR_M})$ contains all Diophantine vectors (actually the inclusion is strict, though it is cumbersome to make the condition $(\mathrm{BR_M})$ explicit in this case).

To conclude, let us give another example in which the growth is so wild that no vectors satisfy the condition $(\mathrm{BR_M})$.

EXAMPLE 6. Let

$$\mu_l = \exp(l^{1/2}), \quad M_l \sim \exp(l^{3/2}).$$

Here too, (H1) and (H2) are satisfied but it is easy to compute

$$C(\sigma) \asymp \exp\left(\frac{1}{4\sigma}\right), \quad C^{-1}(y) \asymp \frac{1}{4\ln y}.$$

Again, this sequence is not matching. Recalling that $\Delta^{-1}(x)$ grows at most like $x^{1/n}$ when x goes to infinity (this is an easy consequence of Dirichlet's box principle), one verifies that the integral in $(\mathrm{BR_M})$ is always divergent, and hence the set $(\mathrm{BR_M})$ is empty.

5. KAM type results

Now let us come back to the Hamiltonian H as in $(*)$, which we recall is given by

$$\begin{cases} H(\theta, I) = h(I) + \varepsilon f(\theta, I), & (\theta, I) \in \mathbb{T}^n \times D \\ \nabla h(0) := \omega_0 \in \mathbb{R}^n. \end{cases}$$

The integrable Hamiltonian h is said to be *non-degenerate* at $0 \in D$ if the Hessian matrix $\nabla^2 h(0) \in M_n(\mathbb{R})$ has a non-zero determinant. The image of the map $\Theta_0 : \mathbb{T}^n \to \mathbb{T}^n \times B$, $\theta \mapsto (\theta, 0)$, is an embedded torus invariant by the flow of h carrying a quasi-periodic flow with frequency ω_0. We shall prove that this quasi-periodic invariant embedded torus is preserved by an arbitrary small perturbation, provided h is non-degenerate, H is M-ultra-differentiable with M satisfying (H1) and (H2), and ω_0 satisfies the $\mathrm{BR_M}$-condition.

THEOREM A. Let H be as in $(*)$, where H is (M, s)-ultra-differentiable with M satisfying (H1) and (H2), $\omega_0 \in \mathbb{R}^n$ satisfying $(\mathrm{BR_M})$ and h is non-degenerate. Then there exists $0 < s' < s$ such that for all ε small enough, there exists an (M, s')-ultra-differentiable torus embedding $\Theta_{\omega_0} : \mathbb{T}^n \to \mathbb{T}^n \times D$ such that $\Theta_{\omega_0}(\mathbb{T}^n)$ is invariant by the Hamiltonian flow of H and quasi-periodic with frequency ω_0. Moreover, Θ_{ω_0} is close to Θ_0 in the sense that

$$|\Theta_{\omega_0} - \Theta_0|_{M, s'} \leq c\sqrt{\varepsilon}$$

for some constant $c \geq 1$ independent of ε.

Theorem A is deduced from a KAM theorem for a Hamiltonian with parameters, for which a quantitative statement is given in § 1. In the analytic case $M = M_1$, this is a result of Rüssmann [**Rüs01**] and in the Gevrey case $M = M_\alpha$, $\alpha \geq 1$, this is a result we proved in [**BF17**]. Theorem A is more general, and it applies for instance to the family $M_{\alpha,\beta}$ given in Example 4 or to the more rapidly growing sequence described in Example 5. As Example 6 shows, if the sequence M grows too fast, the above result is actually void; in the sequel when we will refer to Theorem A we will make the implicit assumption that the $\mathrm{BR_M}$-condition is non-empty.

According to Theorem A, the $\mathrm{BR_M}$-condition is sufficient for the preservation of an invariant torus under an M-ultra-differentiable perturbation. A natural question is: is it necessary? To this question, here we only bring a partial answer. Following Bessi [**Bes00**], one can show that if $\omega = \omega_0$ satisfies a condition (the condition $(\mathrm{B_M})$, which is defined below), the torus can be destroyed. In general, this condition seems unrelated to the $\mathrm{BR_M}$-condition. In the case of matching sequences $M = (M_l)_{l \in \mathbb{N}}$ one can compare them easily. In the case of moderate growth sequences M one can also compare the two conditions, albeit more loosely, since the sequence M may then be upper and lower bounded by matching sequences, as shown by Lemma 46. Recall that associated to M is a function Ω which was defined in (7), and that by definition M is matching if

$$C^{-1}(y) \asymp 1/\Omega(y), \quad y \sim +\infty.$$

THEOREM B. Assume that the vector $\omega \in \mathbb{R}^n$ satisfies the following condition: there exists $c > 0$ such that

$$(\mathrm{B_M}) \qquad \qquad \limsup_{Q \to +\infty} \frac{\ln(\Psi_\omega(Q))}{\Omega(cQ)} > 0.$$

Then an invariant torus with frequency ω can be destroyed by an arbitrary small M-ultra-differentiable perturbation.

Thus the condition that ω_0 does not satisfy $(\mathrm{B_M})$, namely

$$(\mathrm{R_M}) \qquad \qquad \forall\, c > 0, \quad \lim_{Q \to +\infty} \frac{\ln(\Psi_\omega(Q))}{\Omega(cQ)} = 0,$$

is a necessary condition for Theorem A to hold true. Actually, we expect that this condition $(\mathrm{R_M})$ is a necessary and sufficient condition to solve the cohomological equation associated to ω in the M-ultra-differentiable case (in the case $M = M_1$, that this condition is sufficient was shown in [**Rüs75**]).

Now in the case where M is matching, condition $(\mathrm{R_M})$ is equivalent to

$$(13) \qquad \qquad \forall\, c > 0, \quad \lim_{Q \to +\infty} C^{-1}(cQ) \ln(\Psi_\omega(Q)).$$

It is not hard to check that in general, condition $(\mathrm{BR_M})$ implies (13) but clearly they are not equivalent to one another, so, in the matching case, there remains a gap between the sufficient and necessary conditions.

Looking at the family M_α, the necessary condition $(\mathrm{R_M})$ reads

$$\lim_{Q \to +\infty} \frac{\ln(\Psi_\omega(Q))}{Q^{1/\alpha}} = 0$$

which shows that one cannot improve the value of the exponent in the sufficient condition

$$\int_1^{+\infty} \frac{\ln(\Psi_{\omega_0}(Q))}{Q^{1+1/\alpha}} dQ < \infty.$$

More generally, looking at the family $M_{\alpha,\beta}$, $\alpha \geq 1$, $\beta \geq 0$, the necessary condition ($\mathrm{R_M}$) reads

$$\lim_{Q \to +\infty} \frac{\ln(Q)^{\beta/\alpha} \ln(\Psi_\omega(Q))}{Q^{1/\alpha}} = 0,$$

which also shows that one cannot improve the values of the exponents in the sufficient condition

$$\int_{\Delta(1)}^{+\infty} \frac{(\ln(\Delta^{-1}(x)))^{\beta/\alpha}}{x(\Delta^{-1}(x))^{1/\alpha}} dx < \infty.$$

So the gap between the condition ($\mathrm{BR_M}$) and condition ($\mathrm{R_M}$) is narrow; in turn, this means that Theorem A is quite accurate if one restricts attention to matching sequences. For non-matching sequences, Theorem A still gives a non-trivial result (see the Example 5), albeit not necessarily accurate.

We also have variants of Theorem A in the case where h is iso-energetically non-degenerate or the perturbation depends periodically on time.

We say that the integrable Hamiltonian h is *iso-energetically non-degenerate* at 0 if the so-called bordered Hessian of h,

$$\begin{pmatrix} \nabla^2 h(0) & {}^t\nabla h(0) \\ \nabla h(0) & 0 \end{pmatrix},$$

has a non-zero determinant. Under this assumption, the unperturbed torus $I = 0$, with energy $h(0)$, can be continued to a torus with the same energy but with a frequency of the form $\lambda\omega_0$ for λ close to one.

THEOREM C. Let H be as in ($*$), where H is (M, s)-ultra-differentiable with M satisfying (H1) and (H2), $\omega_0 \in \mathbb{R}^n$ satisfying ($\mathrm{BR_M}$) and h is iso-energetically non-degenerate. Then there exists $0 < s' < s$ such that for all ε small enough, there exist $\lambda \in \mathbb{R}^*$ and an (M, s')-ultra-differentiable torus embedding $\Theta_{\omega_0} : \mathbb{T}^n \to \mathbb{T}^n \times D$ such that $\Theta_{\omega_0}(\mathbb{T}^n)$ is invariant by the Hamiltonian flow of H, contained in $H^{-1}(h(0))$ and quasi-periodic with frequency $\lambda\omega_0$. Moreover, λ is close to one and Θ_{ω_0} is close to Θ_0 in the sense that

$$|\lambda - 1| \leq c\sqrt{\varepsilon}, \quad |\Theta_{\omega_0} - \Theta_0|_{M,s'} \leq c\sqrt{\varepsilon}$$

for some constant $c \geq 1$ independent of ε.

We can also look at the non-autonomous time-periodic version; we consider a slightly different setting by looking at a Hamiltonian function $\tilde{H} : \mathbb{T}^n \times D \times \mathbb{T} \to \mathbb{R}$ of the form

($\tilde{*}$) $$\begin{cases} \tilde{H}(\theta, I) = h(I) + \epsilon f(\theta, I, t), \\ \nabla h(0) := \omega_0 \in \mathbb{R}^n. \end{cases}$$

It is better to consider the unperturbed torus $I = 0$ as an invariant torus for the integrable Hamiltonian $\tilde{h} : D \times \mathbb{R} \to \mathbb{R}$ defined by $\tilde{h}(I, J) := h(I) + J$: it is then quasi-periodic with frequency $\tilde{\omega}_0 := (\omega_0, 1)$, has dimension $n + 1$ and is the image of the trivial embedding $\tilde{\Theta}_0 : \mathbb{T}^n \times \mathbb{T} \to \mathbb{T}^n \times D \times \mathbb{T}$.

THEOREM D. Let $\tilde{H}$ be as in ($\tilde{*}$), where $\tilde{H}$ is (M, s)-ultra-differentiable with M satisfying (H1) and (H2), $\omega_0 \in \mathbb{R}^n$ satisfying ($\mathrm{BR_M}$) and h is non-degenerate. Then there exists $0 < s' < s$ such that for all ε small enough, there exists an (M, s')-ultra-differentiable torus embedding $\tilde{\Theta}_{\omega_0} : \mathbb{T}^n \times \mathbb{T} \to \mathbb{T}^n \times D \times \mathbb{T}$ such that $\tilde{\Theta}_{\omega_0}(\mathbb{T}^n \times \mathbb{T})$ is invariant by the Hamiltonian flow of $\tilde{H}$ and quasi-periodic with frequency $\tilde{\omega}_0$. Moreover, $\tilde{\Theta}_{\omega_0}$ is close to $\tilde{\Theta}_0$ in the sense that

$$|\tilde{\Theta}_{\omega_0} - \tilde{\Theta}_0|_{M,s'} \leq c\sqrt{\varepsilon}$$

for some constant $c \geq 1$ independent of ε.

Theorem C and Theorem D are essentially equivalent statements and can be easily deduced from Theorem A; in the analytic case details are given in [**TZ10**], Chapter 2, but it is plain to observe that the arguments still work in the ultra-differentiable case.

We can also state a discrete analogue of Theorem A. Given a function $h : D \to \mathbb{R}$, we define the exact-symplectic map

$$F_h : \mathbb{T}^n \times D \to \mathbb{T}^n \times D, \quad (\theta, I) \mapsto (\theta + \nabla h(I), I).$$

THEOREM E. Let $F : \mathbb{T}^n \times D \to \mathbb{T}^n \times D$ be an (M, s)-ultra-differentiable exact symplectic map, with M satisfying (H1) and (H2) and such that

$$|F - F_h|_{\alpha, s_0} \leq \varepsilon.$$

Assume that $\omega_0 = \nabla h(0)$ satisfies (BR_M) and that h is non-degenerate. Then there exists $0 < s' < s$ such that for all ε small enough, there exists an (M, s')-ultra-differentiable torus embedding $\Theta_{\omega_0} : \mathbb{T}^n \to \mathbb{T}^n \times D$ such that $\Theta_{\omega_0}(\mathbb{T}^n)$ is invariant by F and Θ_{ω_0} gives a conjugacy between the translation of vector ω_0 on $\mathbb{T}^n$ and the restriction of F to $\Theta_{\omega_0}(\mathbb{T}^n)$. Moreover, Θ_{ω_0} is close to Θ_0 in the sense that

$$|\Theta_{\omega_0} - \Theta_0|_{M, s'} \leq c\sqrt{\epsilon}$$

for some constant $c \geq 1$ independent of ε.

Let us recall that Theorem A follows from a KAM theorem for a Hamiltonian with parameters; from this last theorem we will also deduce Arnold's normal form theorem for vector fields on the torus close to constant, in the ultra-differentiable setting.

THEOREM F. Let M be a sequence satisfying (H1) and (H2), $\omega_0 \in \mathbb{R}^n$ satisfying (BR_M) and $X \in \mathcal{U}_{M,s}(\mathbb{T}^n, \mathbb{R}^n)$ a vector field on $\mathbb{T}^n$ of the form

$$X = \omega_0 + B, \quad |B|_{M,s} \leq \mu.$$

Then, for μ sufficiently small, there exist a vector $\omega_0^* \in \mathbb{R}^n$ and an $(M, s/2)$-ultra-differentiable diffeomorphism $\Xi : \mathbb{T}^n \to \mathbb{T}^n$ such that $X + \omega_0^* - \omega_0$ is conjugate to ω_0 via Ξ:

$$\Xi^*(X + \omega_0^* - \omega_0) = \omega_0.$$

Moreover, we have the estimate

$$|\omega_0^* - \omega_0| \leq c\mu, \quad |\Xi - \mathrm{Id}|_{M, s/2} \leq c\mu$$

for some constant $c \geq 1$ independent of μ.

Observe that because of the shift of frequency $\omega_0^* - \omega_0$, in general this result does not give any information on the vector field X. Under some further assumption (for instance, if ω_0 belongs to the rotation set of X, see [**Kar16**]), then this shift vanishes and Theorem F implies that X is conjugated to ω_0.

An even more restricted setting is when X is proportional to ω_0 (so that the flow of X is a re-parametrization of the linear flow of frequency ω_0 and thus ω_0 is the unique rotation vector of X); Theorem F applies in this case to give a conjugacy to ω_0, assuming that ω_0 satisfies (BR_M), but the proof is actually much simpler in this case (it boils down to solve only once the cohomological equation) and should require the weaker condition that ω_0 satisfies (R_M), as it is stated in the analytic case $M = M_1$ in [**Fay02**]. Still in [**Fay02**], it is proved that for $M = M_1$, if ω_0 satisfies (B_M), then there is a dense set of reparametrized linear flow which are weak-mixing (and so cannot be conjugated to the linear flow); thus a necessary condition for Theorem F to hold true is that ω_0 satisfies (R_M) (and this is also a sufficient condition if we impose that X is proportional to ω_0). Clearly, this should extend to the general case of a matching sequence and thus (R_M) is a necessary condition for Theorem F to hold true, as in Theorem A.

Finally, to conclude this section, let us a give the discrete version of Theorem F. Given $\omega_0 \in \mathbb{R}^n$, let T_{ω_0} be the translation of $\mathbb{T}^n$ of vector ω_0:

$$T_{\omega_0} : \mathbb{T}^n \to \mathbb{T}^n, \quad \theta \mapsto \theta + \omega_0.$$

THEOREM G. Let M be a sequence satisfying (H1) and (H2), $\omega_0 \in \mathbb{R}^n$ satisfying (BR_M) and $T \in \mathcal{U}_{M,s}(\mathbb{T}^n, \mathbb{T}^n)$ a diffeomorphism of $\mathbb{T}^n$ of the form

$$T = T_{\omega_0} + B, \quad |B|_{M,s} \leq \mu.$$

Then, for μ sufficiently small, there exist a vector $\omega_0^* \in \mathbb{R}^n$ and an $(M, s/2)$-ultra-differentiable diffeomorphism $\Xi : \mathbb{T}^n \to \mathbb{T}^n$ such that $T + \omega_0^* - \omega_0$ is conjugate to T_{ω_0} via Ξ:

$$\Xi^{-1} \circ (T + \omega_0^* - \omega_0) \circ \Xi = T_{\omega_0}.$$

Moreover, we have the estimate

$$|\omega_0^* - \omega_0| \le c\mu, \quad |\Xi - \mathrm{Id}|_{M,s/2} \le c\mu$$

for some constant $c \ge 1$ independent of μ.

6. Hamiltonian normal forms and Nekhoroshev type results

We are still considering a Hamiltonian H as in $(*)$ but we write ω instead of ω_0, that is

$$\begin{cases} H(\theta, I) = h(I) + \varepsilon f(\theta, I), \quad (\theta, I) \in \mathbb{T}^n \times D \\ \nabla h(0) := \omega_0 = \omega \in \mathbb{R}^n \setminus \{0\}. \end{cases}$$

We do not assume that ω is non-resonant, but we assume at least it is non-zero; without loss of generality (up to a linear symplectic change of coordinates) it can always be written as

$$\omega = (\bar{\omega}, 0) \in \mathbb{R}^d \times \mathbb{R}^{n-d}$$

for some $1 \le d \le n$ and where $\bar{\omega} \in \mathbb{R}^d$ is non-resonant. One can still define Ψ_ω, Δ_ω and Δ_ω^* by

$$\Psi_\omega(Q) := \Psi_{\bar{\omega}}(Q), \quad \Delta_\omega(Q) := Q\Psi_\omega(Q), \quad Q \ge 1$$

and

$$\Delta_\omega^*(x) := \sup\{Q \ge 1 \mid \Delta_\omega(Q) \le x\}, \quad x \ge \Psi_\omega(1).$$

Alternatively, one may also use their continuous and equivalent versions Ψ, Δ and Δ^{-1}.

From now on, let us denote by L_ω the linear integrable Hamiltonian of constant frequency ω, that is $L_\omega(I) = \omega \cdot I$. Our first result is a normal form statement up to a small remainder, in the special case where the integrable Hamiltonian is linear.

THEOREM H. *Let H be as in $(*)$, where H is (M, s)-ultra-differentiable with M satisfying (H1) and (H2) and $h = L_\omega$ with $\omega = (\bar{\omega}, 0) \in \mathbb{R}^d \times \mathbb{R}^{n-d}$ and $\bar{\omega}$ non-resonant. For all ε sufficiently small, there exists a $(M, s/2)$-ultra-differentiable symplectic transformation*

$$\Phi : \mathbb{T}^n \times D_{1/2} \to \mathbb{T}^n \times D$$

such that

$$H \circ \Phi = h + \bar{f} + \hat{f}$$

where $\{\bar{f}, h\} = 0$ and with the estimates

$$\begin{cases} |\Phi - \mathrm{Id}|_{s/2} \le c_1 \Psi(\Delta^{-1}(c_2 s \varepsilon^{-1}))\varepsilon, \quad |\bar{f}|_{s/2} \le c_1 \varepsilon, \\ |\hat{f}|_{s/2} \le c_1 \varepsilon \exp\left(-c_3 \left(C^{-1}(c_4 s \Delta^{-1}(c_2 s \varepsilon^{-1}))\right)^{-1}\right) \end{cases}$$

where $c_1 > 1$, $c_2 < 1$, $c_3 < 1$ and $c_4 < 1$ depend only on n, ω and the function C.

This result was known in the analytic case (see [**Fas90**], [**Pös93**]) and in the Gevrey case ([**Bou13a**]). Theorem H implies the following "partial" stability result, in which we denote by $\Pi_d : \mathbb{R}^n \to \mathbb{R}^d$ the projection onto the first d components.

COROLLARY 1. *Under the assumptions of Theorem H, for any $I_0 \in D_{1/8}$ and any $r > 0$ such that*

$$2c_1 \Psi(\Delta^{-1}(c_2 s \varepsilon^{-1}))\varepsilon \le r \le 1/4,$$

any solution $(\theta(t), I(t))$ of H with $I(0) = I_0$ satisfy

$$|\Pi_d(I(t) - I_0)| \le 2r, \quad |t| \le \tilde{c}_1 r s \varepsilon^{-1} \exp\left(-c_3 \left(C^{-1}(c_4 s \Delta^{-1}(c_2 s \varepsilon^{-1}))\right)^{-1}\right)$$

as long as $I(t) \in D_{1/4}$, where $c_1 > 1$, $\tilde{c}_1 < 1$ $c_2 < 1$, $c_3 < 1$ and $c_4 < 1$ depend only on n, ω and the function C.

Let us point out that these stability estimates actually apply to any solution; indeed, if $I(0) \in D_{1-\delta}$ for an arbitrary but fixed $\delta > 0$, one would obtain

$$|\Pi_d(I(t) - I(0))| < \delta$$

for the same interval of time as long as $I(t)$ stays in the domain, upon letting the constants and the smallness assumption depend on δ. For convenience only, we sated the result for any solution with $I(0) \in D_{1/8}$ and the restriction on time was that $I(t)$ stays in $D_{1/4}$. In general, this result does not give a confinement of the action variables for a long interval of time because it may well happen that the orbit escapes the domain very quickly; the only exception is when $d = n$, that is when ω is actually non-resonant. The next corollary follows at once from the preceding one.

COROLLARY 2. *Under the assumptions of Theorem H and in the case where $\omega = \bar{\omega} \in \mathbb{R}^n$ is non-resonant, for any $I_0 \in D_{1/8}$ and any $r > 0$ such that*

$$2c_1 \Psi(\Delta^{-1}(c_2 s \varepsilon^{-1}))\varepsilon \leq r \leq 1/4,$$

any solution $(\theta(t), I(t))$ of H with $I(0) = I_0$ satisfy

$$|(I(t) - I_0)| \leq 2r, \quad |t| \leq \tilde{c}_1 r s \varepsilon^{-1} \exp\left(-c_3 \left(C^{-1}(c_4 s \Delta^{-1}(c_2 s \varepsilon^{-1}))\right)^{-1}\right)$$

where $c_1 > 1$, $\tilde{c}_1 < 1$ $c_2 < 1$, $c_3 < 1$ and $c_4 < 1$ depend only on n, ω and the function C.

A natural question is whether these last estimates can be improved in general. Without any assumption on the sequence $M = (M_l)_{l \in \mathbb{N}}$, we can always construct an example of unstable orbits with controlled speed in the special case where the integrable Hamiltonian h is linear. In the statement below, for convenience we will measure the size of the perturbation f not by its norm but rather by the norm of its Hamiltonian vector field X_f.

THEOREM I. *There exists a sequence of Hamiltonians $H_j = h + f_j = L_\omega + f_j$ as in $(*)$ where $\omega = (\bar{\omega}, 0) \in \mathbb{R}^d \times \mathbb{R}^{n-d}$ with $\bar{\omega}$ non-resonant and*

$$|X_{f_j}|_s \leq 2c\varepsilon_j, \quad \lim_{j \to +\infty} \varepsilon_j = 0,$$

such that H_j has a solution $(\theta_j(t), I_j(t))$ which is globally defined and satisfies

$$|t|\varepsilon_j \exp\left(-\Omega(16\pi s \Delta_\omega^*(2\varepsilon_j^{-1}))\right) \leq |\Pi_d(I_j(t) - I_j(0))| \leq |t|\varepsilon_j \exp\left(-\Omega(8\pi s \Delta_\omega^*((2\varepsilon_j)^{-1}))\right).$$

In the analytic case, this theorem was proved in [**Bou12**] and then extended in [**Bou13a**] in the Gevrey case.

In general, this last statement is not related to Theorem H or Corollary 1 because it is the function Ω, and not C^{-1}, which measures the speed of instability. Yet for matching sequences, for which by definition $1/\Omega$ is equivalent to C^{-1}, Theorem I shows that in general one cannot improve the conclusions of both Corollary 1 and Theorem H.

Theorem H gives a global normal form for a linear integrable Hamiltonian $h = L_\omega$ and Corollary 1 gives a global stability result. For an arbitrary nonlinear integrable Hamiltonian h, these global results translate into local results, valid on a small ball of radius $\rho > 0$ around the origin. The only results we can obtain have an anistropic character, so for a given $\rho > 0$, it is convenient to introduce the scalings

$$\sigma_\rho(\theta, I) := (\theta, \rho I), \quad \sigma_\rho^{-1}(\theta, I) := (\theta, \rho^{-1}I).$$

Clearly, σ_ρ is a diffeomorphism between $\mathbb{T}^n \times D$ and $\mathbb{T}^n \times D_\rho$.

THEOREM J. *Let H be as in $(*)$, where H is (M, s)-ultra-differentiable with M satisfying (H1) and (H2) and $\omega = (\bar{\omega}, 0) \in \mathbb{R}^d \times \mathbb{R}^{n-d}$ with $\bar{\omega}$ non-resonant. For all ε and all ρ sufficiently small with*

$$\sqrt{\varepsilon} \leq \rho \leq 1,$$

there exists a $(M, s/4)$-ultra-differentiable symplectic transformation

$$\Phi : \mathbb{T}^n \times D_{\rho/2} \to \mathbb{T}^n \times D_\rho$$

such that

$$H \circ \Phi = h + \bar{f} + \hat{f}$$

where $\{\bar{f}, L_\omega\} = 0$ and with the estimates

$$\begin{cases} |\sigma_\rho^{-1} \circ \Phi \circ \sigma_\rho - \mathrm{Id}|_{s/4} \le c_1 \Psi(\Delta^{-1}(c_2 s \rho^{-1})) \rho^{-1} \varepsilon, & |\bar{f} \circ \sigma_\rho|_{s/4} \le c_1 \varepsilon, \\ |\hat{f} \circ \sigma_\rho|_{s/4} \le c_1 \varepsilon \exp\left(-c_3 \left(C^{-1}(c_4 s \Delta^{-1}(c_2 s \rho^{-1}))\right)^{-1}\right) \end{cases}$$

where $c_1 > 1$, $c_2 < 1$, $c_3 < 1$ and $c_4 < 1$ depend only on n, ω, $|h|_s$ and the function C.

In the analytic or Gevrey case, a similar result but with weaker estimates was obtained in [**Bou13a**]. As before, Theorem J implies the following "partial" and local stability result.

COROLLARY 3. *Under the assumptions of Theorem J, for any $I_0 \in D_{\rho/8}$ and any $r > 0$ such that*

$$2c_1 \Psi(\Delta^{-1}(c_2 s \varepsilon^{-1})) \varepsilon \le r \le \rho/4,$$

any solution $(\theta(t), I(t))$ of H with $I(0) = I_0$ satisfy

$$|I(t) - I_0| \le 2r, \quad |t| \le \tilde{c}_1 r s \varepsilon^{-1} \exp\left(-c_3 \left(C^{-1}(c_4 s \Delta^{-1}(c_2 s \rho^{-1}))\right)^{-1}\right)$$

as long as $I(t) \in D_{\rho/4}$, where $c_1 > 1$, $\tilde{c}_1 < 1$ $c_2 < 1$, $c_3 < 1$ and $c_4 < 1$ depend only on n, ω, $|h|_s$ and the function C.

Again, we also have the special case where the frequency is non-resonant.

COROLLARY 4. *Under the assumptions of Theorem J and in the case where $\omega = \bar{\omega} \in \mathbb{R}^n$ is non-resonant, for any $I_0 \in D_{\rho/8}$ and any $r > 0$ such that*

$$2c_1 \Psi(\Delta^{-1}(c_2 s \varepsilon^{-1})) \varepsilon \le r \le \rho/4,$$

any solution $(\theta(t), I(t))$ of H with $I(0) = I_0$ satisfy

$$|I(t) - I_0| \le 2r, \quad |t| \le \tilde{c}_1 r s \varepsilon^{-1} \exp\left(-c_3 \left(C^{-1}(c_4 s \Delta^{-1}(c_2 s \rho^{-1}))\right)^{-1}\right)$$

where $c_1 > 1$, $\tilde{c}_1 < 1$ $c_2 < 1$, $c_3 < 1$ and $c_4 < 1$ depend only on n, ω, $|h|_s$ and the function C.

For an arbitrary non-linear integrable Hamiltonian, one can only obtain local and partial stability result; moreover we do not have an equivalent of Theorem I as Corollary 3 can be improved in many cases.

In order to obtain global and full stability, one needs a further assumption on the integrable part. Let us say that $h : G \to \mathbb{R}$ is *quasi-convex* on some domain $G \subseteq \mathbb{R}^n$ if there exist positive constants l and m, such that for all $I \in G$, it satisfies the following two properties:

$$(14) \qquad\qquad\qquad |\nabla h(I)| \ge l$$

and at least of the inequalities

$$(15) \qquad\qquad |\nabla h(I) \cdot \xi| \ge l|\xi|, \quad |\nabla^2 h(I) \xi \cdot \xi| \ge m|\xi|^2$$

is satisfied for any $\xi \in \mathbb{R}^n$. To emphasize the role of the constants l and m, we say that h is (l, m)-quasi-convex. This definition is slightly more restrictive than the one used in [**Pös93**], as (14) is not assumed in the latter reference; we actually require h not to have critical points just for convenience, in order to have a statement which is uniform in phase space (this is not the case in [**Pös93**]). Observe that if we assume (14) and only the second part of (15) for all $\xi \in \mathbb{R}^n$ orthogonal to $\nabla h(I)$, then (15) is actually satisfied for a possibly smaller constant l (see [**MS02**] for instance).

Under this quasi-convexity assumption, we will prove the following result.

THEOREM K. *Let H be as in $(*)$, where H is (M, s)-ultra-differentiable with M satisfying (H1) and (H2) and h is (l, m)-quasi-convex. For all ε sufficiently small, any $I_0 \in D_{1/2}$ and any solution $(\theta(t), I(t))$ of H with $I(0) = I_0$, we have*

$$|I(t) - I_0| \le c_1 s (s^{-2} \varepsilon)^{\frac{1}{2n}}, \quad |t| \le \tilde{c}_1 s \exp\left(c_2 \left(C^{-1}\left(c_3 s (s^2 \varepsilon^{-1})^{\frac{1}{2n}}\right)\right)^{-1}\right)$$

where $c_1 > 1$, $c_2 < 1$ and $c_3 < 1$ depend only on n, $|h|_s$, l, m and the function C.

In the analytic case $M = M_1$, this theorem is due to Nekhoroshev ([**Nek77**],[**Nek79**]) with an exponent worse than $1/(2n)$; the acutal value $1/(2n)$ was later obtained in [**Loc92**], [**LN92**], [**LNN94**] and [**Pös93**]. The Gevrey case $M = M_\alpha$ for $\alpha \geq 1$ was established in [**MS02**], and so our Theorem K further extend these results for any sequence M satisfying (H1) and (H2).

As before, the natural question is whether these stability estimates can be improved in general, and to do so, one has to construct unstable orbits. As a matter of fact, the exponent $1/(2n)$ in the time-scale of stability in Theorem K can be improved if one is willing to allow a drift arbitrarily close to one (that is, replacing the drift in the action variables of order $\varepsilon^{1/(2n)}$ by a drift of order ε^b with b arbitrarily close to zero). In the Gevrey case $M = M_\alpha$ for $\alpha \geq 1$ (thus including the analytic case), it was proved in [**BM11**] that one can get an exponent arbitrarily close to $1/(2(n-1))$ and in the analytic case, it was proved in [**ZZ17**] that one can get an exponent arbitrarily close to $1/(2(n-2))$.

For analytic Hamiltonians, this ends the question of optimality as examples of unstable orbits with exponent $1/(2(n-2))$ where constructed in [**Bes96**] for $n = 3$, [**Bes97**] for $n = 4$ and [**Zha11**] for $n \geq 5$. For Gevrey non-analytic Hamiltonians $\alpha > 1$ a similar result was obtained in [**MS02**], but the latter crucially use the assumption that the space of α-Gevrey functions contains bump functions.

For a given sequence $M = (M_l)_{l\in\mathbb{N}}$, it is said to be non-quasi-analytic if the following condition is satisfied:

(H3)   The sequence $(1/\mu_l)_{l\in\mathbb{N}} = (M_l/M_{l+1})_{l\in\mathbb{N}}$ is summable, $i.e., \sum_{l\in\mathbb{N}} M_l/M_{l+1} < +\infty.$

Due to a famous theorem of Denjoy-Carleman, this condition exactly characterizes classes which are non-quasi-analytic; we recall here that a function (say, on the real line) is quasi-analytic if it vanishes identically if and only if its formal Taylor series vanishes at some point. Actually, non-quasi-analytic classes contain bump functions (with arbitrarily small support) and this follows from the proof of Denjoy-Carleman theorem.

As it turns out, we are not able to use the mechanism of [**Bes96**], [**Bes97**] and [**Zha11**] but we can still use the construction of [**MS02**], for a sequence M which satisfy (H1), (H2) but also (H3). Here is a precise statement.

THEOREM L. Let $n \geq 3$. There exists a sequence of Hamiltonians $H_j = h + f_j$ as in $(*)$, where H_j is (M, s)-ultra-differentiable with M satisfying (H1), (H2) and (H3), h satisfies (14) and (15) and

$$|f_j|_s \leq \varepsilon_j, \quad \lim_{j\to+\infty} \varepsilon_j = 0,$$

such that the following holds. The Hamiltonian H_j has a solution $(\theta_j(t), I_j(t))$ which satisfies

$$|I_j(\tau_j) - I_j(0)| \geq 1$$

where $\tau_j > 0$ satisfies, for j large enough, the estimates

$$\exp\left(2\Omega\left(c_1\varepsilon_j^{-\frac{1}{2(n-2)}}\right)\right) \leq \tau_j \leq \exp\left(2\Omega\left(c_2\varepsilon_j^{-\frac{1}{2(n-2)}}\right)\right)$$

for positive constants $c_1 < c_2 < 1$ which depend only n, s and the sequence M.

Leaving aside the question of the exponent $1/(2(n-2))$ that we already discussed, Theorem L says that Theorem K is essentially optimal, at least for matching sequences which satisfy (H3).

To conclude, we recall that Nekhoroshev actually proved Theorem K in the analytic case not only for quasi-convex functions, but for a much larger and generic class of functions that we now define.

We say that a differentiable function $h : G \to \mathbb{R}$ is *steep* on some domain $G \subseteq \mathbb{R}^n$ if there exist positive constants l, L, δ, p_j, for any integer $1 \leq j \leq n-1$, such that for all $I \in G$, we have $|\nabla h(I)| \geq l$ and, for all integer $1 \leq j \leq n-1$, for all vector space $\Lambda \in \mathbb{R}^n$ of dimension j, letting

$\lambda = I + \Lambda$ the associated affine subspace passing through I and h_λ the restriction of h to λ, the inequality

$$\max_{0 \leq \eta \leq \xi} \min_{|I' - I| = \eta, \, I' \in \lambda} |\nabla h_\lambda(I') - \nabla h_\lambda(I)| > L\xi^{p_j}$$

holds true for all $0 < \xi \leq \delta$. As before, to emphasize the role of the constants we say that h is $(l, L, \delta, (p_j)_{1 \leq j \leq n-1})$-steep on G and, if all the $p_j = p$, we say that h is (l, L, δ, p)-steep on G.

One may easily check that quasi-convex functions are steep, and in fact they are the "steepest" as one has $p = 1$ in this case (in general p_j are integers larger than one). But steep functions form a much larger class of functions and among sufficiently smooth functions they are generic in many senses (see [**Nek79**] or [**BFN17**]).

The original proof of Nekhoroshev, in the analytic case, was dealing with steep integrable Hamiltonians but the values of the stability exponents, which depend on the steepness exponents p_j for $1 \leq j \leq n$, were not very sharp and even restricted to the quasi-convex they would be of order $1/n^2$ which is obviously worse than the exponent $1/(2n)$ of Theorem K. A different and somehow simpler proof was given in [**BN12**], using only rough periodic approximations, but leading also to bad values for the exponents; the interest was that this method was flexible enough to give results in the Gevrey (and even finitely differentiable) case, see [**Bou11**]. Sharp exponents, generalizing those of the quasi-convex case, were finally obtained in [**GCB16**] but so far it is restricted to the analytic case. So in the more general setting of ultra-differentiable functions we are considering here, we were not able to extend the results of [**GCB16**], but we will manage to use the method originally introduced in [**BN12**] to obtain the following result.

THEOREM M. *Let H be as in $(*)$, where H is (M, s)-ultra-differentiable with M satisfying (H1) and (H2) and h is (l, L, δ, p)-steep. For all ε sufficiently small, any $I_0 \in D_{1/2}$ and any solution $(\theta(t), I(t))$ of H with $I(0) = I_0$, we have*

$$|I(t) - I_0| \leq c_1 \varepsilon^{\frac{1}{2na}}, \quad |t| \leq s \exp\left(c_2 \left(C^{-1} \left(c_3 s(\varepsilon^{-1})^{\frac{1}{2na}} \right) \right)^{-1} \right)$$

where $a := (np)^{n-1}$ and $c_1 > 1$, $c_2 < 1$ and $c_3 < 1$ which depend only on n, $|h|_s$, l, L, δ, p and the function C.

As we already explained before, the value of the exponent $1/(2na)$ is very bad here and one should expect better exponents such as those obtained in [**GCB16**]. However unlike the quasi-convex Theorem K which comes with Theorem L to test the optimality of the exponents, we do not have such instability result in the steep non-quasi-convex case; this is actually an open problem in any regularity, including the analytic one.

Estimates on ultra-differentiable functions

Let us start by recalling some notations and definitions. Given an integer $m \geq 1$ and $k = (k_1, \ldots, k_m) \in \mathbb{N}^m$, we define

$$|k| = \sum_{i=1}^{m} k_i, \quad k! = \prod_{i=1}^{m} k_i!.$$

Given $x \in \mathbb{R}^m$, we set

$$x^k = \prod_{i=1}^{m} x_i^{k_i}$$

and for a smooth function $f : B \to \mathbb{R}$ defined on some open ball $B \subset \mathbb{R}^m$ (see Remark 1 at the end of § 1 for functions defined on an open set of $\mathbb{R}^m \times \mathbb{T}^n$) and $a \in B$, we set

$$\partial^k f(a) = \partial_{x_1}^{k_1} \cdots \partial_{x_m}^{k_m} f(a).$$

Given a sequence $M = (M_l)_{l \in \mathbb{N}}$ with $M_0 = M_1 = 1$ and $s > 0$, recall from § 2 that the function f is said to be (M, s)-ultra-differentiable if

(16)
$$|f|_{M,s} := c \sup_{a \in B} \left(\sup_{k \in \mathbb{N}^m} \frac{(|k| + 1)^2 s^{|k|} |\partial^k f(a)|}{M_{|k|}} \right) < \infty, \quad c := 4\pi^2/3.$$

The space of such functions will be denoted by $\mathcal{U}_{M,s}(B)$, and, equipped with the above norm, it is a Banach space. On the one hand, the role of the factor $(|k|+1)^2$ is to make easier the estimates for the product and for the composition of ultra-differentiable functions (see respectively Lemmas 11 and 14). On the other hand, the role of the normalizing constant $c > 0$ in the definition is to ensure that $\mathcal{U}_{M,s}(B)$ is a Banach algebra with respect to multiplication (see Lemma 11). The factor $(|k|+1)^2$ is only here for convenience and it can be removed (see § 4 and § 5); then one has to choose a different normalizing constant to turn $\mathcal{U}_{M,s}(B)$ into a Banach algebra.

The above definition can be extended to vector-valued functions $f = (f_i)_{1 \leq i \leq p} : K \to \mathbb{R}^p$ for $p \geq 1$ by setting

$$|\partial^k f(a)| := \sum_{1 \leq i \leq p} |\partial^k f_i(a)|, \quad a \in K$$

in (16). The space of such vector-valued functions is still a Banach space with the above norm, and it will be denoted by $\mathcal{U}_{M,s}(B, \mathbb{R}^p)$. The case of matrix-valued functions, say with values in the space $M_{m,p}(\mathbb{R})$ of matrices with m rows and p columns, is reduced to the case of vector-valued functions by simply identifying $M_{m,p}(\mathbb{R})$ to $\mathbb{R}^{mp}$.

In the sequel, to lighten the notations, we will simply write $|\,.\,|_s$ (respectively $\mathcal{U}_s(B)$ and $\mathcal{U}_s(B, \mathbb{R}^p)$) instead of $|\,.\,|_{M,s}$ (respectively $\mathcal{U}_{M,s}(B)$ and $\mathcal{U}_{M,s}(B, \mathbb{R}^p)$).

1. Majorant series and ultra-differentiable functions

The definition of ultra-differentiable functions can be conveniently reformulated in terms of majorant series with one variable (see [**Kom79**], [**Kom80**] and also [**SCK03**]).

But first let us consider a formal power series in m variables $X = (X_1, \ldots, X_m)$ with coefficients in a normed real vector space $(E, |\,.\,|_E)$, which is a formal sum of the form

$$A(X) = \sum_{k \in \mathbb{N}^m} A_k X^k, \quad A_k = A_{k_1, \ldots, k_m} \in E.$$

Such a formal series is said to be majorized by another formal power series with real non-negative coefficients

$$B(X) = \sum_{k \in \mathbb{N}^m} B_k X^k, \quad B_k = B_{k_1, \ldots, k_m} \in \mathbb{R}^+,$$

and we write $A \ll B$, if

$$(17) \qquad\qquad\qquad |A_k|_E \leq B_k, \quad \forall\, k \in \mathbb{N}^m.$$

Next, following [**SCK03**], we introduce a notion of a smooth function being majorized by a formal power series in one variable. So let $f : B \to \mathbb{R}^p$ be a smooth function defined on some open ball $B \subset \mathbb{R}^m$, and F be a formal power series in one variable with non-negative coefficients, that we shall write as

$$F(X) = \sum_{l=0}^{+\infty} \frac{F_l}{l!} X^l.$$

We will say that f is majorized by F on B, and we will write $f \ll_B F$ (or $f(x) \ll_B F(X)$), if for all $a \in B$ and all $k \in \mathbb{N}^m$, we have

$$(18) \qquad\qquad\qquad |\partial^k f(a)| \leq F_{|k|}.$$

It will be usefull to rephrase this relation in terms of the formal Taylor series of f at points $a \in B$ (taking values in $\mathbb{R}^p$), defined by

$$T_a f(X) := \sum_{k \in \mathbb{N}^m} \frac{\partial^k f(a)}{k!} X^k.$$

To the formal series F in one variable, one can associate a formal series $\hat{F}$ in m variables simply by setting

$$\hat{F}(X_1, \ldots, X_m) := F(X_1 + \cdots + X_m).$$

Then

$$f \ll_B F \quad \text{in the sense of (18)} \Leftrightarrow \forall a \in B \quad T_a f \ll \hat{F} \quad \text{in the sense of (17)}$$

(with $E = \mathbb{R}^p$ and $|\,.\,|_E = |\,.\,|_{l^1}$).

Now, given a sequence M and $s > 0$, let us define the following formal power series in one variable

$$(19) \qquad U_s(X) := c^{-1} \sum_{l=0}^{+\infty} \frac{1}{(l+1)^2} \frac{M_l}{l!} \left(\frac{X}{s} \right)^l = c^{-1} \sum_{l=0}^{+\infty} \frac{N_l}{(l+1)^2} \left(\frac{X}{s} \right)^l, \quad c = 4\pi^2/3$$

(recall from (4) that we have defined $N_l := M_l/l!$). The following characterization of ultra-differentiable functions is evident from the definitions (16) and (18).

PROPOSITION 5. *If $f : B \to \mathbb{R}^p$ is a smooth function and $s > 0$, then*

$$|f|_s = \inf \{ C \in [0, +\infty] \mid f \ll_B C U_s \}$$

and

$$f \in \mathcal{U}_s(B, \mathbb{R}^p) \Leftrightarrow |f|_s < \infty.$$

Remark 1. Note that being an ultra-differentiable function is a local property. So the above immediately translates for functions defined on an open subet of the product of $\mathbb{T}^n$ and $\mathbb{R}^m$, by merely considering any local chart at any point of the torus. These definitions thus encompass the space $\mathcal{U}_s(\mathbb{T}^n \times D, \mathbb{R}^p)$ introduced in § 2.

2. Properties of majorant series

We collect here some properties of majorant series that will be used later on. It is clear how to define the derivatives of a formal power series in one variable, and also a linear combination and the product of two such formal power series. We then have the following lemma.

LEMMA 6. *Let $f, g : B \to \mathbb{R}^p$ be smooth functions, F, G be two formal power series in one variable, and assume that*

$$f \ll_B F, \quad g \ll_B G.$$

Then

$$(20) \qquad \qquad \partial^k f \ll_B \partial^{|k|} F, \quad k \in \mathbb{N}^m,$$

$$(21) \qquad \qquad \lambda f + \mu g \ll_B |\lambda| F + |\mu| G, \quad \lambda \in \mathbb{R}, \ \mu \in \mathbb{R}.$$

Moreover, the scalar product $f \cdot g = \sum_{i=1}^p f_i g_i : B \to \mathbb{R}$ satisfies

$$(22) \qquad \qquad f \cdot g \ll_B FG.$$

For a proof, we refer to [**SCK03**], Lemma 2.2, in which the case $p = 1$ is considered; but the general case $p \geq 1$ is entirely similar.

Given two formal power series in one variable F and G, we define the composition $F \odot G$ of F and G by

$$F \odot G(X) := \sum_{l=0}^{+\infty} \frac{F_l}{l!} \left(G(X) - G(0) \right)^l.$$

LEMMA 7. *Let $f : B \to \mathbb{R}^p$, $g : D \to \mathbb{R}^m$ be smooth functions where B and D are open balls of $\mathbb{R}^m$ such that $g(D) \subseteq B$, and assume that*

$$f \ll_B F, \quad g \ll_D G.$$

Then

$$f \circ g \ll_D F \odot G.$$

Once again, for a proof we refer to [**SCK03**], Lemma 2.3.

3. Derivatives

In this section, we will show that derivatives of an M-ultra-differentiable function are still M-ultra-differentiable, at the expense of reducing the width parameter $s > 0$; these are analogues of Cauchy estimates for analytic functions. It is precisely at this point that the assumption (H2) on the sequence M is needed and that the Cauchy function C comes into play.

PROPOSITION 8. *Let $f \in \mathcal{U}_s(B, \mathbb{R}^p)$ and $0 < \sigma < 1$. Then for any $k \in \mathbb{N}^m$ with $|k| = 1$, $\partial^k f \in G_{s(1-\sigma)}(B, \mathbb{R}^p)$ and we have*

$$|\partial^k f|_{s(1-\sigma)} \leq s^{-1} C(\sigma) |f|_s.$$

We have stated the above proposition only for $k \in \mathbb{N}^m$ with $|k| = 1$ as this is the only case we shall need; but clearly one could obtain an estimate for any $k \in \mathbb{N}^m$ by a straightforward induction.

PROOF. From Proposition 5 and (20) of Lemma 6, it is sufficient to prove that

$$(23) \qquad \qquad \partial^1 U_s \ll s^{-1} C(\sigma) U_{s(1-\sigma)}$$

where U_s is the formal power series defined in (19). We have

$$\partial^1 U_s(X) = c^{-1} \sum_{l=1}^{+\infty} \frac{1}{(l+1)^2} \frac{M_l}{(l-1)!} \frac{X^{l-1}}{s^l} = c^{-1} \sum_{l=0}^{+\infty} \frac{1}{(l+2)^2} \frac{M_{l+1}}{l!} \frac{X^l}{s^{l+1}}$$

and hence (23) is true if, for all $l \in \mathbb{N}$,

$$\frac{M_{l+1}}{M_l} (1-\sigma)^l \leq C(\sigma)$$

which is trivially true by the definition of the function C in (5) and since $1 - \sigma \leq e^{-\sigma}$ for $0 \leq \sigma \leq 1$.   □

For $f : B \to \mathbb{R}$, let $\nabla f : B \to \mathbb{R}^m$ be the vector-valued function formed by the partial derivatives of f of order one, and more generally, for $f : B \to \mathbb{R}^p$, we let $\nabla f : B \to M_{m,p}(\mathbb{R}) \simeq \mathbb{R}^{mp}$ be the matrix-valued function whose columns are given by ∇f_i where $f = (f_i)_{1 \leq i \leq p}$. Then we have the following obvious corollary of Proposition 8.

COROLLARY 9. *Let $0 < \sigma < 1$. If $f \in \mathcal{U}_s(B, \mathbb{R})$, then $\nabla f \in \mathcal{U}_{s(1-\sigma)}(B, \mathbb{R}^m)$ and*

$$|\nabla f|_{s(1-\sigma)} \leq ms^{-1}C(\sigma)|f|_s$$

and if $f \in \mathcal{U}_s(B, \mathbb{R}^p)$, then $\nabla f \in \mathcal{U}_{s(1-\sigma)}(B, \mathbb{R}^{mp})$ and

$$|\nabla f|_{s(1-\sigma)} \leq mps^{-1}C(\sigma)|f|_s.$$

4. Products

In this section, we shall prove that the space of M-ultra-differentiable functions is stable under multiplication, provided (H1) is satisfied.

PROPOSITION 10. *Let $f, g \in \mathcal{U}_s(B, \mathbb{R}^p)$, with M satisfying* (H1). *Then $f \cdot g \in \mathcal{U}_s(B, \mathbb{R})$ and we have*

$$|f \cdot g|_s \leq |f|_s |g|_s.$$

Once again, in view of Proposition 5 and (22) of Proposition 6, Proposition 10 is a direct consequence of the following lemma.

LEMMA 11. *We have*

$$U_s^2 \ll U_s.$$

To prove this lemma, we will need another elementary lemma which makes use of the condition (H1).

LEMMA 12. *If M satisfies* (H1), *then*

$$N_l N_k \leq N_{l+k}, \quad (l, k) \in \mathbb{N}^2.$$

PROOF. Since the sequence ν is increasing and $N_0 = 1$, we have

$$N_l N_k = \nu_0 \ldots \nu_{l-1} \nu_0 \ldots \nu_{k-1} \leq \nu_0 \ldots \nu_{l-1} \nu_l \ldots \nu_{l+k-1} = N_{l+k}.$$

                                                                                                □

The proof of Lemma 11 given below follows [**Lax53**]. Observe that we only need the conclusion of Lemma 12, which is thus weaker than the condition (H1). It is this latter lemma that motivates the introduction of the normalizing constant in U_s (and thus in the ultra-differentiable norm); without this constant one would have $U_s^2 \ll cU_s$. Let us point out that the proof given below is elementary thanks to the factor $(|k| + 1)^2$ in the definition of the norm (16) (which gives the factor $(l + 1)^2$ in the defintion of U_s); without this factor, the statement is true (with a different normalizing constant) but the proof is more involved (see Lemma 2.7 of [**SCK03**] for the sequence M_α, $\alpha \geq 1$).

PROOF OF LEMMA 11. Recall that

$$U_s(X) = c^{-1} \sum_{l=0}^{+\infty} \frac{1}{(l+1)^2} \frac{M_l}{l!} \left(\frac{X}{s}\right)^l = c^{-1} \sum_{l=0}^{+\infty} \frac{N_l}{(l+1)^2} \left(\frac{X}{s}\right)^l$$

and so the assertion of the lemma amounts to prove that for all $l \in \mathbb{N}$,

$$(24) \qquad \sum_{j=0}^{l} \frac{N_j N_{l-j}}{(j+1)^2 (l-j+1)^2} \leq c \frac{N_l}{(l+1)^2}, \quad c = \frac{4\pi^2}{3}.$$

Observe that the sum in the left-hand side of (24) is symmetric with respect to $j \mapsto l - j$, and that from Lemma 12, $N_j N_{l-j} \leq N_l$ for all $l \in \mathbb{N}$ and all $0 \leq j \leq l$. Hence,

$$\sum_{j=0}^{l} \frac{N_j N_{l-j}}{(j+1)^2 (l-j+1)^2} \leq 2 \sum_{j=0}^{l/2} \frac{N_j N_{l-j}}{(j+1)^2 (l-j+1)^2} \leq 2 N_l \sum_{j=0}^{l/2} \frac{1}{(j+1)^2 (l-j+1)^2}.$$

Then for any $0 \leq j \leq l/2$, $(l-j+1)^2 \geq (l/2+1)^2 \geq (l+1)^2/4$, and therefore

$$\sum_{j=0}^{l} \frac{N_j N_{l-j}}{(j+1)^2 (l-j+1)^2} \leq \frac{8 N_l}{(l+1)^2} \sum_{j=0}^{l/2} \frac{1}{(j+1)^2} \leq \frac{8 N_l}{(l+1)^2} \sum_{j=0}^{+\infty} \frac{1}{(j+1)^2} = \frac{4\pi^2}{3} \frac{N_l}{(l+1)^2}$$

which is the inequality we wanted to prove. $\square$

Proposition 10 can be extended to matrix-valued functions. More precisely, given $f : B \to M_{m,p}(\mathbb{R})$ and $g : B \to M_{p,q}(\mathbb{R})$ where $f = (f_{i,j})_{1 \leq i \leq m, \, 1 \leq j \leq p}$ and $g = (g_{j,k})_{1 \leq j \leq p, \, 1 \leq k \leq q}$, we define $f \cdot g : B \to M_{m,q}(\mathbb{R})$ by $f \cdot g := ((f \cdot g)_{i,k})_{1 \leq i \leq m, \, 1 \leq k \leq q}$ where

$$(f \cdot g)_{i,k} := \sum_{j=1}^{p} f_{i,j} g_{j,k}.$$

Then the following statement is an obvious corollary of Proposition 10.

COROLLARY 13. *Let* $f \in \mathcal{U}_s(B, M_{m,p}(\mathbb{R}))$, $g \in \mathcal{U}_s(B, M_{p,q}(\mathbb{R}))$. *Then* $f \cdot g \in \mathcal{U}_s(B, M_{m,q}(\mathbb{R}))$ *and we have*

$$|f \cdot g|_s \leq |f|_s |g|_s.$$

5. Compositions

Our goal here is to prove that the space of M-ultra-differentiable functions is stable under composition, provided M satisfies (H1), at the expense of reducing the width $s > 0$ as with derivatives. The main point is that we will be able to reduce the width by a factor arbitrarily close to one provided the composition involves a map arbitrarily close to the identity.

But first we need to define two additional formal power series associated to U_s, the latter being defined in (19). So we define $\bar{U}_s$ by

$$(25) \qquad \qquad \bar{U}_s(X) := U_s(X) - U_s(0) = c^{-1} \sum_{l=1}^{+\infty} \frac{N_l}{(l+1)^2} \left(\frac{X}{s} \right)^l$$

and $\tilde{U}_s$ by

$$(26) \qquad \qquad \tilde{U}_s(X) := c^{-1} \sum_{l=0}^{+\infty} \frac{N_{l+1}}{(l+2)^2} \left(\frac{X}{s} \right)^l.$$

It is clear that

$$(27) \qquad \qquad s^{-1} X \tilde{U}_s(X) = \bar{U}_s(X).$$

LEMMA 14. *We have*

$$\tilde{U}_s^2 \ll \tilde{U}_s, \quad \tilde{U}_s \bar{U}_s \ll \bar{U}_s.$$

As for Lemma 11, the factor $(l+1)^2$ in the definition of U_s makes the proof simple, but the statement is still true without this factor (see Lemma 2.4 of [**SCK03**] for the sequence M_α, $\alpha \geq 1$). To prove Lemma 14, we shall need the following preliminary lemma, in which we use the assumption (H1).

LEMMA 15. *If* M *satisfies* (H1)*, then*

$$N_l N_k \leq N_{l+k-1}, \quad (l,k) \in \mathbb{N}^2 \setminus \{0\}.$$

PROOF. We assume, without loss of generality, that $l \geq 1$. Since the sequence ν is increasing and $N_0 = 1$, we have

$$N_l N_k = \nu_0 \ldots \nu_{l-1} \nu_0 \nu_1 \ldots \nu_{k-1} \leq \nu_0 \ldots \nu_{l-1} \nu_0 \nu_l \ldots \nu_{l+k-2} = \nu_0 N_{l+k-1}$$

and the result follows since we normalized $\nu_0 = N_1/N_0 = 1$. □

Observe that exactly as for Lemma 11, we only need the conclusion of Lemma 15 to hold true, which is thus weaker than the condition (H1).

PROOF OF LEMMA 14. It is enough to prove the first part of the statement, as the second part of the statement follows from it; indeed, if $\tilde{U}_s^2 \ll \tilde{U}_s$, then using (27) we have

$$\tilde{U}_s(X)\bar{U}_s(X) = s^{-1} X \tilde{U}_s(X) \tilde{U}_s(X) \ll s^{-1} X \tilde{U}_s(X) = \bar{U}_s(X).$$

As in Lemma 11, to prove that $\tilde{U}_s^2 \ll \tilde{U}_s$ one needs to show

$$(28) \qquad \sum_{j=0}^{l} \frac{N_{j+1} N_{l-j+1}}{(j+2)^2 (l-j+2)^2} \leq c \frac{N_{l+1}}{(l+2)^2}, \quad c = \frac{4\pi^2}{3}.$$

The sum in the left-hand side of (28) is still symmetric with respect to $j \mapsto l - j$, and for any $l \in \mathbb{N}$ and any $0 \leq j \leq l$, we have $N_{j+1} N_{l-j+1} \leq N_{l+1}$ by Lemma 15. Therefore

$$\sum_{j=0}^{l} \frac{N_{j+1} N_{l-j+1}}{(j+2)^2 (l-j+2)^2} \leq 2 \sum_{j=0}^{l/2} \frac{N_{j+1} N_{l-j+1}}{(j+2)^2 (l-j+2)^2} \leq 2 N_{l+1} \sum_{j=0}^{l/2} \frac{1}{(j+2)^2 (l-j+2)^2}.$$

Moreover, as in Lemma 11, for any $0 \leq j \leq l/2$, $(l-j+2)^2 \geq (l/2+2)^2 \geq (l+2)^2/4$, and so the last inequality leads to

$$\sum_{j=0}^{l} \frac{N_{j+1} N_{l-j+1}}{(j+2)^2 (l-j+2)^2} \leq \frac{8 N_{l+1}}{(l+2)^2} \sum_{j=0}^{l/2} \frac{1}{(j+2)^2} \leq \frac{4\pi^2}{3} \frac{N_{l+1}}{(l+2)^2}$$

which concludes the proof. □

PROPOSITION 16. Let $f \in \mathcal{U}_s(B, \mathbb{R}^p)$, $0 < \sigma < 1$, $g \in \mathcal{U}_{s(1-\sigma)}(D, \mathbb{R}^p)$ where B and D are open balls of $\mathbb{R}^m$ such that $g(D) \subseteq B$. If

$$(29) \qquad |g - \mathrm{Id}|_{s(1-\sigma)} \leq s\sigma,$$

then $f \circ g \in \mathcal{U}_{s(1-\sigma)}(D, \mathbb{R}^p)$ and

$$|f \circ g|_{s(1-\sigma)} \leq |f|_s.$$

As it will be clear in the proof, the conclusions of Proposition 16 hold true under the slightly weaker assumption that

$$|g - \mathrm{Id}|_{s(1-\sigma)} - \sup_{x \in D} |g(x) - x| \leq s\sigma$$

but this will not be needed.

PROOF. Let $u = g - \mathrm{Id} \in \mathcal{U}_{s(1-\sigma)}(D, \mathbb{R}^p)$ and

$$a := |f|_s, \quad b := |u|_{s(1-\sigma)}$$

so that, from Proposition 5,

$$f(x) \ll_B a U_s(X), \quad u(x) \ll_D b U_{s(1-\sigma)}(X)$$

and consequently

$$f(x) \ll_B a U_s(X), \quad g(x) \ll_D X + b U_{s(1-\sigma)}(X).$$

We now apply Lemma 7 and, recalling the definition of $\bar{U}_s$ and $\tilde{U}_s$ given respectively in (25) and (26), we obtain

$$
\begin{aligned}
f(g(x)) \quad \ll_B \quad & aU_s\left(X + b\bar{U}_{s(1-\sigma)}(X)\right) \\
= \quad & aU_s(0) + a\bar{U}_s\left(X + b\bar{U}_{s(1-\sigma)}(X)\right) \\
= \quad & aU_s(0) + a\bar{U}_s\left(X + b(s(1-\sigma))^{-1}X\tilde{U}_{s(1-\sigma)}(X)\right) \\
= \quad & aU_s(0) + a\sum_{l=1}^{+\infty} \frac{N_l}{(l+1)^2}\left(\frac{X + b(s(1-\sigma))^{-1}X\tilde{U}_{s(1-\sigma)}(X)}{s}\right)^l \\
= \quad & aU_s(0) + a\sum_{l=1}^{+\infty} \frac{N_l}{(l+1)^2}\left(\frac{X}{s}\right)^l \left(1 + b(s(1-\sigma))^{-1}\tilde{U}_{s(1-\sigma)}(X)\right)^l.
\end{aligned}
$$

(30)

From the first part of Lemma 14, for any $j \in \mathbb{N}$, we have

$$
\tilde{U}_{s(1-\sigma)}^j \ll \tilde{U}_{s(1-\sigma)}
$$

and therefore

$$
\begin{aligned}
\left(1 + b(s(1-\sigma))^{-1}\tilde{U}_{s(1-\sigma)}(X)\right)^l \quad = \quad & \sum_{j=0}^{l}\binom{l}{j}b^j(s(1-\sigma))^{-j}\tilde{U}_{s(1-\sigma)}(X)^j \\
\ll \quad & \tilde{U}_{s(1-\sigma)}(X)\sum_{j=0}^{l}\binom{l}{j}b^j(s(1-\sigma))^{-j} \\
= \quad & \tilde{U}_{s(1-\sigma)}(X)\left(1 + b(s(1-\sigma))^{-1}\right)^l.
\end{aligned}
$$

Now, from (29) we get

$$
b = |u|_{s(1-\sigma)} \le s\sigma
$$

and thus

$$
1 + b(s(1-\sigma))^{-1} \le 1 + \sigma(1-\sigma)^{-1} = (1-\sigma)^{-1}.
$$

This gives

$$
\left(1 + b(s(1-\sigma))^{-1}\tilde{U}_{s(1-\sigma)}(X)\right)^l \ll \tilde{U}_{s(1-\sigma)}(X)(1-\sigma)^{-l}
$$

which, together with (30), yields

$$
\begin{aligned}
f(g(x)) \quad \ll_D \quad & aU_s(0) + a\tilde{U}_{s(1-\sigma)}(X)\sum_{l=1}^{+\infty}\frac{N_l}{(l+1)^2}\left(\frac{X}{(s(1-\sigma))}\right)^l \\
= \quad & aU_s(0) + a\tilde{U}_{s(1-\sigma)}(X)\bar{U}_{s(1-\sigma)}(X).
\end{aligned}
$$

Using the second part of Lemma 14, this gives

$$
f(g(x)) \ll_D aU_s(0) + a\bar{U}_{s(1-\sigma)}(X)
$$

and since $U_s(0) = U_{s(1-\sigma)}(0)$, we arrive at

$$
f(g(x)) \ll_D a(U_{s(1-\sigma)}(0) + \bar{U}_{s(1-\sigma)}(X)) = aU_{s(1-\sigma)}(X).
$$

Using Proposition 5, we eventually obtain

$$
|f \circ g|_{s(1-\sigma)} \le a = |f|_s
$$

and this concludes the proof. $\qquad\square$

6. Flows

The goal of this section is to show that the flow of an M-ultra-differentiable vector field is M-ultra-differentiable, under the assumptions (H1) and (H2), and to obtain a norm estimate on this flow in terms of the norm of the vector field; to prove this we will use the estimates on derivatives, products and composition we already obtained in § 3, § 4 and § 5 and the classical contraction fixed point theorem.

Here we will consider a situation adapted to the applications in § 2 and § 3; that is we consider functions $H = H(\theta, I, \omega)$ which are defined and ultra-differentiable on a domain of the form

$$\mathbb{T}^n \times D_r \times D_h(\omega_0) \subseteq \mathbb{T}^n \times \mathbb{R}^n \times \mathbb{R}^n$$

where D_r is the open ball of radius $r > 0$ centered at the origin and $D_h(\omega_0)$ is the open ball of radius $h > 0$ centered at a vector $\omega_0 \in \mathbb{R}^n$. In the statements below, the variables $\omega \in D_h(\omega_0)$ play the role of a fixed parameter, hence to simplify the notations we will explicitly suppress the dependence on $\omega \in D_h(\omega_0)$.

Let us first start with a vector-valued function $D : \mathbb{T}^n \to \mathbb{R}^n$ which depends only on $\theta \in \mathbb{T}^n$, and that we shall consider as a vector field on $\mathbb{T}^n$.

LEMMA 17. *Given $D \in \mathcal{U}_s(\mathbb{T}^n, \mathbb{R}^n)$, let $0 < \sigma < 1$ and assume that*

$$(31) \qquad\qquad |D|_s \leq s\sigma.$$

Then for any $t \in [0, 1]$, the time-t map D^t of the flow of D belongs to $\mathcal{U}_{s(1-\sigma)}(\mathbb{T}^n, \mathbb{T}^n)$ and we have the estimate

$$(32) \qquad\qquad |D^t - \mathrm{Id}|_{s(1-\sigma)^2} \leq t|D|_s.$$

Observe that the width decreases from s to $s(1-\sigma)^2$ because we will have to apply consecutively Propositions 16 and 8 (or, more precisely, Corollary 9), and each application induces a decrease of s by a factor $1 - \sigma$.

PROOF. The fact that D^t is smooth and defined for all $t \in [0, 1]$ (in fact, for all $t \in \mathbb{R}$) follows from the compactness of $\mathbb{T}^n$ and the classical result on the existence and uniqueness of solutions of differential equations (even though this will essentially be re-proved below); the only thing we need to prove is the estimate (32), and clearly it is sufficient to prove that for all $t \in [0, 1]$,

$$|D^t - \mathrm{Id}|_{s(1-\sigma)^2} \leq |D|_s.$$

So let us consider the space $V := C([0, 1], \mathcal{U}_{s(1-\sigma)^2}(\mathbb{T}^n, \mathbb{T}^n))$ of continuous maps from $[0, 1]$ to $\mathcal{U}_{s(1-\sigma)^2}(\mathbb{T}^n, \mathbb{T}^n)$: given an element $\Phi \in V$ and $t \in [0, 1]$, we shall write $\Phi^t := \Phi(t)$ and consequently $\Phi = (\Phi^t)_{t \in [0,1]}$. We equip V with the following norm:

$$||\Phi|| := \max_{t \in [0,1]} |\Phi^t|_{s(1-\sigma)^2}$$

which makes it a Banach space, and if we set $\rho := |D|_s$, we define

$$B_\rho V := \{\Phi \in V \mid ||\Phi - \mathrm{Id}|| \leq \rho\}.$$

We can eventually define a Picard operator P associated to D by

$$P : B_\rho V \to V, \quad \Phi \mapsto P(\Phi)$$

where $P(\Phi) = (P(\Phi)^t)_{t \in [0,1]}$ is defined by

$$P(\Phi)^t := \mathrm{Id} + \int_0^t D \circ \Phi^\tau d\tau.$$

To prove the lemma, it is sufficient to prove that P has a unique fixed point $\Phi_* \in B_\rho V$, as necessarily $(\Phi_*^t)_{t \in [0,1]} = (D^t)_{t \in [0,1]}$. Therefore it is sufficient to prove that P sends $B_\rho V$ into itself and that one of its iterate is a contraction, as $B_\rho V$ is a complete subset of the Banach space V.

First we need to show that P maps $B_\rho V$ into itself. So assume $\Phi \in B_\rho V$, using (31) this implies that for all $t \in [0, 1]$,

$$|\Phi^t - \mathrm{Id}|_{s(1-\sigma)^2} \leq \rho \leq s\sigma.$$

Letting σ^* be such that $(1-\sigma)^2 = 1 - \sigma^*$, we have $\sigma < \sigma^*$ and therefore

$$|\Phi^t - \mathrm{Id}|_{s(1-\sigma^*)} = |\Phi^t - \mathrm{Id}|_{s(1-\sigma)^2} \le s\sigma \le s\sigma^*$$

so that (29) of Proposition 16 is satisfied (with $f = D$ and $g = \Phi^t$ for any $t \in [0,1]$) and the latter proposition applies: this gives

$$|D \circ \Phi^t|_{s(1-\sigma)^2} = |D \circ \Phi^t|_{s(1-\sigma^*)} \le |D|_s = \rho, \quad t \in [0,1]$$

hence

$$\left| P(\Phi)^t - \mathrm{Id} \right|_{s(1-\sigma)^2} = \left| \int_0^t D \circ \Phi^\tau d\tau \right|_{s(1-\sigma)^2} \le t\rho \le \rho, \quad t \in [0,1]$$

and therefore

$$\| P(\Phi) - \mathrm{Id} \| \le \rho.$$

This proves that P maps $B_\rho V$ into itself.

It remains to show that some iterate of P is a contraction. So let $\Phi_1, \Phi_2 \in B_\rho V$, then for any $t \in [0,1]$,

$$
\begin{aligned}
P(\Phi_1)^t - P(\Phi_2)^t &= \int_0^t (D \circ \Phi_1^\tau - D \circ \Phi_2^\tau)\, d\tau \\
&= \int_0^t \left(\int_0^1 \nabla D \circ (s\Phi_1^\tau + (1-s)\Phi_2^\tau) ds \right) \cdot (\Phi_1^\tau - \Phi_2^\tau) d\tau \\
&= \int_0^t Y^\tau \cdot (\Phi_1^\tau - \Phi_2^\tau) d\tau.
\end{aligned}
$$

Using (29), we can apply Proposition 16, and together with Corollary 9, we obtain, for any $0 \le \tau \le t$,

$$|Y^\tau|_{s(1-\sigma)^2} \le |\nabla D|_{s(1-\sigma)} \le n^2 s^{-1} C(\sigma)|D|_s.$$

Using Corollary 13, this gives, for any $t \in [0,1]$,

$$|P(\Phi_1)^t - P(\Phi_2)^t|_{s(1-\sigma)^2} \le t n^2 s^{-1} C(\sigma)|D|_s \max_{0 \le \tau \le t} |\Phi_1^\tau - \Phi_2^\tau|_{s(1-\sigma)^2}$$

and hence, setting $\kappa := n^2 s^{-1} C(\sigma)|D|_s$,

$$\max_{0 \le \tau \le t} |P(\Phi_1)^\tau - P(\Phi_2)^\tau|_{s(1-\sigma)^2} \le \kappa t \max_{0 \le \tau \le t} |\Phi_1^\tau - \Phi_2^\tau|_{s(1-\sigma)^2}$$

A well-known induction then gives, for any $j \in \mathbb{N}$,

$$\max_{0 \le \tau \le t} |P^j(\Phi_1)^\tau - P^j(\Phi_2)^\tau|_{s(1-\sigma)^2} \le \frac{(\kappa t)^j}{j!} \max_{0 \le \tau \le t} |\Phi_1^\tau - \Phi_2^\tau|_{s(1-\sigma)^2}$$

and hence setting $t = 1$

$$\| P^j(\Phi_1) - P^j(\Phi_2) \| \le \frac{\kappa^j}{j!} \| \Phi_1 - \Phi_2 \|.$$

This proves that some iterate of P is a contraction, which concludes the proof. $\qquad \square$

Now let us consider a Hamiltonian function X on $\mathbb{T}^n \times D_r$, of the form

(33) $$X(\theta, I) := C(\theta) + D(\theta) \cdot I, \quad C : \mathbb{T}^n \to \mathbb{R}, \quad D : \mathbb{T}^n \to \mathbb{R}^n.$$

The Hamiltonian equations associated to X are given by:

$$
\begin{cases}
\dot\theta(t) = \nabla_I X(\theta(t), I(t)) = D(\theta(t)), \\
\dot I(t) = -\nabla_\theta X(\theta(t), I(t)) = -\nabla C(\theta(t)) - \nabla D(\theta(t)) \cdot I.
\end{cases}
$$

The equations for θ are uncoupled from the equations of I (and hence can be integrated independently), while the equations for I are affine in I; it is well-known that these facts lead to a simple form of the Hamiltonian flow associated to X (see, for instance, [**Vil08**]).

PROPOSITION 18. *Let X be as in* (33) *with $C \in \mathcal{U}_s(\mathbb{T}^n, \mathbb{R})$ and $D \in \mathcal{U}_s(\mathbb{T}^n, \mathbb{R}^n)$. Let $0 < \sigma < 1$ and assume that*

$$(34) \qquad\qquad |D|_s \leq s\sigma.$$

Then for any $t \in [0, 1]$, the time-t map X^t of the Hamiltonian flow of X is of the form

$$X^t(\theta, I) = (\theta + E^t(\theta), I + F^t(\theta) \cdot I + G^t(\theta))$$

where $E^t \in \mathcal{U}_{s(1-\sigma)^2}(\mathbb{T}^n, \mathbb{R}^n)$, $F^t \in \mathcal{U}_{s(1-\sigma)^2}(\mathbb{T}^n, \mathbb{R}^{n^2})$ and $G^t \in \mathcal{U}_{s(1-\sigma)^2}(\mathbb{T}^n, \mathbb{R}^n)$ with the estimates

$$(35) \qquad \begin{cases} |E^t|_{s(1-\sigma)^2} \leq t|D|_s, \\ |F^t|_{s(1-\sigma)^2} \leq n^2 s^{-1} C(\sigma) |D|_s \exp(tn^2 s^{-1} C(\sigma)|D|_s), \\ |G^t|_{s(1-\sigma)^2} \leq ns^{-1} C(\sigma)|C|_s \exp(tn^2 s^{-1} C(\sigma)|D|_s). \end{cases}$$

As a consequence, given $0 < \delta < r$, if we further assume that

$$(36) \qquad ns^{-1} C(\sigma) \exp(tn^2 s^{-1} C(\sigma)|D|_s)(rn|D|_s + |C|_s) \leq \delta$$

then X^t maps $\mathbb{T}^n \times D_{r-\delta}$ into $\mathbb{T}^n \times D_r$.

PROOF. The second part of the statement clearly follows from the first part, so let us prove the latter. From the specific form of the Hamiltonian equations associated to X, one has, for any $t \in [0, 1]$,

$$X^t(\theta, I) = (\theta + E^t(\theta), I + F^t(\theta) \cdot I + G^t(\theta))$$

with

$$\begin{cases} E^t(\theta) = \int_0^t D(\theta + E^\tau(\theta))d\tau, \\ F^t(\theta) = -\int_0^t \nabla D(\theta + E^\tau(\theta))d\tau - \int_0^t \nabla D(\theta + E^\tau(\theta)) \cdot F^\tau(\theta)d\tau \\ G^t(\theta) = -\int_0^t \nabla C(\theta + E^\tau(\theta))d\tau - \int_0^t \nabla D(\theta + E^\tau(\theta)) \cdot G^\tau(\theta)d\tau. \end{cases}$$

Because of (34), Lemma 17 applies and the flow $D^t(\theta) = \theta + E^t(\theta)$ satisfies (32), and therefore

$$|E^t|_{s(1-\sigma)^2} = |D^t - \mathrm{Id}|_{s(1-\sigma)^2} \leq t|D|_s$$

which gives the first estimate of (35). Using this estimate and (34), we can apply Proposition 16 and Corollary 9 to obtain, for any $0 \leq \tau \leq t \leq 1$,

$$|\nabla D \circ D^\tau|_{s(1-\sigma)^2} \leq |\nabla D|_{s(1-\sigma)} \leq n^2 s^{-1} C(\sigma)|D|_s.$$

Looking at the expression of F^t, this gives

$$|F^t|_{s(1-\sigma)^2} \leq n^2 s^{-1} C(\sigma)|D|_s \left(1 + \int_0^t |F^\tau|_{s(1-\sigma)^2}d\tau\right)$$

which, by Gronwall's inequality, implies that for all $t \in [0, 1]$,

$$|F^t|_{s(1-\sigma)^2} \leq n^2 s^{-1} C(\sigma)|D|_s \exp(tn^2 s^{-1} C(\sigma)|D|_s)$$

which gives the second estimate of (35). For the third estimate of (35), observe that the same argument yields

$$|G^t|_{s(1-\sigma)^2} \leq ns^{-1} C(\sigma)|C|_s + n^2 s^{-1} C(\sigma)|D|_s \int_0^t |G^\tau|_{s(1-\sigma)^2}d\tau$$

and again, by Gronwall's inequality, for all $t \in [0, 1]$ we have

$$|G^t|_{s(1-\sigma)^2} \leq ns^{-1} C(\sigma)|C|_s \exp(tn^2 s^{-1} C(\sigma)|D|_s).$$

This concludes the proof. □

The last lemma gives estimates on the Hamiltonian flow associated to a Hamiltonian X on $\mathbb{T}^n \times D_r$ which is affine in I; the flow has then a special form and precise estimates can be obtained (they are needed in § 2). On the other hand, in § 3, we will need estimates valid for an arbitrary Hamiltonian Y on $\mathbb{T}^n \times D_r$; the lemma below gives rougher estimates in this general case but they will prove sufficient for our purpose.

PROPOSITION 19. *Given $Y \in \mathcal{U}_s(\mathbb{T}^n \times D_r)$, let $0 < \sigma < 1$ and assume that*

$$(37) \qquad |Y|_s \le (2n)^{-1} s^2 \sigma C(\sigma)^{-1}.$$

Then for any $t \in [0,1]$, the time-t map Y^t of the Hamiltonian flow of Y belongs to $\mathcal{U}_{s(1-\sigma)^3}(\mathbb{T}^n \times D_{r(1-\sigma)^3}, \mathbb{T}^n \times D_{r(1-\sigma)^2})$ and we have the estimate

$$(38) \qquad |Y^t - \mathrm{Id}|_{s(1-\sigma)^3} \le t|Y|_s.$$

PROOF. The proof is just a variation of the proof of Lemma 17, so we just point out the differences. As before, we know that the flow Y^t exists and is smooth, so we only need to prove that Y^t maps $\mathbb{T}^n \times D_{r(1-\sigma)^3}$ into $\mathbb{T}^n \times D_{r(1-\sigma)^2}$ and that the estimate

$$(39) \qquad |Y^t - \mathrm{Id}|_{s(1-\sigma)^3} \le |Y|_s$$

holds true. Let X_Y be the Hamiltonian vector field: it follows from Corollary 9 and (37) that

$$(40) \qquad |X_Y|_{s(1-\sigma)} \le 2ns^{-1}C(\sigma)|Y|_s \le s\sigma$$

and thus Y^t maps $\mathbb{T}^n \times D_{r(1-\sigma)^i}$ into $\mathbb{T}^n \times D_{r(1-\sigma)^{i-1}}$ for $i = 1, 2, 3$. To prove the estimate (39), at this point one can just repeat the proof of Lemma 17, with $s(1-\sigma)$ instead of s, X_Y instead of D; the required condition (31) translates into $|X_Y|_{s(1-\sigma)} \le s\sigma$ which is nothing but (40). $\qquad \square$

7. Inverse functions

In this last section, we shall prove that if an M-ultra-differentiable map is sufficiently close to the identity, then its local inverse is still M-ultra-differentiable, provided M satisfies (H1) and (H2). Exactly like in § 6, we will use the estimates on derivatives, products and composition we obtained in § 3, § 4 and § 5 and the classical contraction fixed point theorem.

Again, to prove this in a setting adapted to § 2, let us consider a map ϕ which depends only on $\omega \in D_h(\omega_0)$, that is $\phi : D_h(\omega_0) \to \mathbb{R}^n$.

PROPOSITION 20. *Given $\phi \in \mathcal{U}_s(D_h(\omega_0), \mathbb{R}^n)$, let $0 < \sigma < 1$ and assume that*

$$(41) \qquad |\phi - \mathrm{Id}|_s < n^{-2} s C(\sigma)^{-1}, \quad |\phi - \mathrm{Id}|_s \le h/2.$$

Then there exists a unique $\varphi \in G_{s(1-\sigma)^2}(D_{h/2}(\omega_0), D_h(\omega_0))$ such that $\phi \circ \varphi = \mathrm{Id}$ and

$$(42) \qquad |\varphi - \mathrm{Id}|_{s(1-\sigma)^2} \le |\phi - \mathrm{Id}|_s.$$

PROOF. Let us define $V := \mathcal{U}_{s(1-\sigma)^2}(D_{h/2}(\omega_0), \mathbb{R}^n)$, which is a Banach space with the norm $||\cdot|| = |\cdot|_{s(1-\sigma)^2}$, and for $\rho := |\phi - \mathrm{Id}|_s$, we set

$$B_\rho V := \{\psi \in V \mid ||\psi - \mathrm{Id}|| \le \rho\}.$$

Let us define the following Picard operator P associated to ϕ:

$$P : B_\rho V \to V, \quad \psi \mapsto P(\psi) = \mathrm{Id} - (\phi - \mathrm{Id}) \circ \psi.$$

It is clear that $\phi \circ \varphi = \mathrm{Id}$ if and only if φ is a fixed point of P, and therefore the proposition will be proved once we have shown that P has a unique fixed point in $B_\rho V$, and to do this it is enough to prove that P is a well-defined contraction of $B_\rho V$.

First let us prove that P maps $B_\rho V$ into itself. So let $\psi \in B_\rho V$, and using the second part of (41), observe that since

$$\sup_{\omega \in D_{h/2}(\omega_0)} |\psi(\omega) - \omega| \le ||\psi - \mathrm{Id}|| \le \rho \le h/2$$

then ψ maps $D_{h/2}(\omega_0)$ into $D_h(\omega_0)$. Now recall that for any $0 < \sigma < 1$, $C(\sigma) \ge (e\sigma)^{-1}$ so the first part of (41) gives in particular

$$||\psi - \mathrm{Id}|| \le \rho \le s\sigma$$

and allows us to apply Proposition 16 (with $\sigma^* > \sigma$ defined by $(1-\sigma)^2 = (1 - \sigma^*)$) to get

$$||(\phi - \mathrm{Id}) \circ \psi|| = |(\phi - \mathrm{Id}) \circ \psi|_{s(1-\sigma)^2} \le |\phi - \mathrm{Id}|_s = \rho$$

and thus

$$||P(\psi) - \mathrm{Id}|| = ||(\phi - \mathrm{Id}) \circ \psi|| \le \rho,$$

that is, P maps $B_\rho V$ into itself. To show that P is a contraction, using Corollary 9, Corollary 13 and Proposition 16 one gets (exactly as in the proof of Lemma 17) for any $\psi_1, \psi_2 \in B_\rho V$:

$$||P(\psi_1) - P(\psi_2)|| = ||(\phi - \text{Id}) \circ \psi_1 - (\phi - \text{Id}) \circ \psi_2|| \leq n^2 s^{-1} C(\sigma) |\phi - \text{Id}|_{\alpha, s} ||\psi_1 - \psi_2||$$

and thus the fact that P is a contraction follows from the first part of (41). This ends the proof. $\quad\square$

CHAPTER 2

Application to KAM theory

In this section, we give applications to KAM theory, namely we prove Theorem A, Theorem F and Theorem B; Theorem A and Theorem F will be deduced from a KAM theorem with parameters (Theorem N below) and Theorem B is an adaptation of a result of [**Bes00**]. All the other KAM theorems we stated can be deduced from Theorem A, Theorem F or Theorem N so we will not give details about their proof.

In the special case where the Hamiltonians are Gevrey regular, that is $M = M_\alpha$ for $\alpha \geq 1$, these results are contained in [**BF17**] and so it is our purpose here to extend them to the more general setting we are considering.

1. Statement of the KAM theorem with parameters

Let us now consider the following setting. Fix $\omega_0 \in \mathbb{R}^n \setminus \{0\}$. Re-ordering the components of ω_0 and re-scaling the Hamiltonian allow us to assume without loss of generality that

$$\omega_0 = (1, \bar{\omega}_0) \in \mathbb{R}^n, \quad \bar{\omega}_0 \in [-1, 1]^{n-1}.$$

Given real numbers $r > 0$ and $h > 0$, we let

$$D_r := \{I \in \mathbb{R}^n \mid |I| < r\}, \quad D_h(\omega_0) := \{\omega \in \mathbb{R}^n \mid |\omega - \omega_0| < h\}.$$

Since ω_0 is fixed, for simplicity we shall remove it from the notation and write $D_h = D_h(\omega_0)$ (somewhat imprecisely), and we set

$$D_{r,h} := D_r \times D_h.$$

Our Hamiltonians will be defined on $\mathbb{T}^n \times D_{r,h}$, a neighborhood of $\mathbb{T}^n \times \{0\} \times \{\omega_0\}$ in $\mathbb{T}^n \times \mathbb{R}^n \times \mathbb{R}^n$.

Let $s > 0$, $\eta \geq 0$ a fixed parameter, $\varepsilon \geq 0$ and $\mu \geq 0$ two small parameters. We consider a function $H \in \mathcal{U}_s(\mathbb{T}^n \times D_{r,h})$ of the form

$$(**) \quad \begin{cases} H(\theta, I, \omega) = \underbrace{e(\omega) + \omega \cdot I}_{N(I,\omega)} + \underbrace{A(\theta, \omega) + B(\theta, \omega) \cdot I}_{P(\theta,I,\omega)} + \underbrace{M(\theta, I, \omega) \cdot I^2}_{R(\theta,I,\omega)} \\ |A|_s \leq \varepsilon, \quad |B|_s \leq \mu, \quad |\nabla_I^2 R|_s \leq \eta \end{cases}$$

where the notation $M(\theta, I, \omega) \cdot I^2$ stands for the vector I given twice as an argument to the symmetric bilinear form $M(\theta, I, \omega)$. Observe that $A : \mathbb{T}^n \times D_h \to \mathbb{R}$, $B : \mathbb{T}^n \times D_h \to \mathbb{R}^n$ whereas $M : \mathbb{T}^n \times D_{r,h} \to M_n(\mathbb{R})$ with $M_n(\mathbb{R})$ the ring of real square matrices of size n.

The function H in $(**)$ should be considered as an ultra-differentiable Hamiltonian on $\mathbb{T}^n \times D_r$, depending on a parameter $\omega \in D_h$; for a fixed parameter $\omega \in D_h$, when convenient, we will write

$$H_\omega(\theta, I) = H(\theta, I, \omega), \quad N_\omega(I) = N(I, \omega), \quad P_\omega(\theta, I) = P(\theta, I, \omega), \quad R_\omega(\theta, I) = R(\theta, I, \omega).$$

The image of the map $\Phi_0 : \mathbb{T}^n \to \mathbb{T}^n \times D_r$, $\theta \mapsto (\theta, 0)$ is an ultra-differentiable embedded torus in $\mathbb{T}^n \times D_r$, invariant by the Hamiltonian flow of $N_{\omega_0} + R_{\omega_0}$ and quasi-periodic with frequency ω_0. The next theorem asserts that this quasi-periodic torus will persist, being only slightly deformed, as an invariant torus not for the Hamiltonian flow of H_{ω_0} but for the Hamiltonian flow of $H_{\omega_0^*}$, where ω_0^* is a parameter close to ω_0, provided ε and μ are sufficiently small and ω_0 satisfies the BR$_M$-condition. Here is the precise statement.

THEOREM N. Let H be as in $(**)$, with $M = (M_l)_{l \in \mathbb{N}}$ satisfying (H1) and (H2), and ω_0 satisfying (BR_M). There exist positive constants $c_1 \leq 1$, $c_2 \leq 1$ and $c_3 \geq 1$ depending only on n such that if

$$(43) \qquad\qquad \sqrt{\varepsilon} \leq \mu \leq h/2, \quad \sqrt{\varepsilon} \leq r, \quad h \leq c_1(Q_0\Psi(Q_0))^{-1}$$

where $Q_0 \geq n + 2$ is sufficiently large so that

$$(44) \qquad C^{-1}(c_2(1+\eta)^{-1}sQ_0) + (\ln 2)^{-1} \int_{\Delta(Q_0)}^{+\infty} C^{-1}(c_2(1+\eta)^{-1}s\Delta^{-1}(x))\frac{dx}{x} \leq \ln 2(4n+2)^{-1}$$

then the following holds true. There exist a vector $\omega_0^* \in \mathbb{R}^n$ and an $(M, s/2)$-ultra-differentiable embedding

$$\Phi_{\omega_0}^* : \mathbb{T}^n \times D_{r/2} \to \mathbb{T}^n \times D_r$$

of the form

$$\Phi_{\omega_0}^*(\theta, I) = (\theta + E^*(\theta), I + F^*(\theta) \cdot I + G^*(\theta))$$

with the estimates

$$(45) \qquad |\omega_0^* - \omega_0| \leq c_3\mu, \quad |E^*|_{s/2} \leq c_3\Psi(Q_0)\mu, \quad |F^*|_{s/2} \leq c_3\Delta(Q_0)\mu, \quad |G^*|_{s/2} \leq c_3\Delta(Q_0)\varepsilon$$

and such that

$$H_{\omega_0^*} \circ \Phi_{\omega_0}^*(\theta, I) = e_0^* + \omega_0 \cdot I + R^*(\theta, I), \quad R^*(\theta, I) = M^*(\theta, I) \cdot I^2,$$

with the estimates

$$|e_0^* - e_{\omega_0^*}| \leq c_3\varepsilon, \quad |\nabla_I^2 R^* - \nabla_I^2 R_{\omega_0^*}|_{s/2} \leq c_3\eta\Delta(Q_0)\mu.$$

Theorem A follows quite directly from Theorem N, introducing the frequencies $\omega = \nabla h(I)$ as independent parameters, taking $\mu = \sqrt{\varepsilon}$, and tuning the shift of frequency $\omega_0^* - \omega_0$ using the non-degeneracy assumption on the unperturbed Hamiltonian. Theorem F follows also from Theorem N by realizing X as the restriction of a Hamiltonian vector field on an invariant torus, setting $\varepsilon = \eta = 0$ and letting μ be the only small parameter. These arguments are made precise in § 5 - § 6, but we first concentrate on the proof of Theorem N.

In all this section, we do not pay attention to how constants depend on the dimension n as it is fixed. Hence in the sequel we shall write

$$u \lessdot v \quad (\text{respectively } u \cdot\!\!\lessdot v, \quad u =\!\!\cdot\, v, \quad u \cdot\!\!= v)$$

if, for some constant $c \geq 1$ depending only on n, we have

$$u \leq cv \quad (\text{respectively } cu \leq v, \quad u = cv, \quad cu = v).$$

2. Approximation by rational vectors

Recall that we have written $\omega_0 = (1, \bar{\omega}_0) \in \mathbb{R}^n$ with $\bar{\omega}_0 \in [-1, 1]^{n-1}$. For a given $Q \geq 1$, it is always possible to find a rational vector $v = (1, p/q) \in \mathbb{Q}^n$, with $p \in \mathbb{Z}^{n-1}$ and $q \in \mathbb{N}$, which is a Q-approximation in the sense that $|q\omega_0 - qv| \leq Q^{-1}$, and for which the denominator q satisfies the upper bound $q \leq Q^{n-1}$: this is essentially the content of Dirichlet's theorem on simultaneous rational approximations, and it holds true without any assumption on ω_0. In our situation, since we have assumed that ω_0 is non-resonant, there exist not only one, but n linearly independent rational vectors in $\mathbb{Q}^n$ which are Q-approximations. Moreover, one can obtain not only linearly independent vectors, but rational vectors $v_1, \ldots, v_n$ of denominators $q_1, \ldots, q_n$ such that the associated integer vectors $q_1v_1, \ldots, q_nv_n$ form a $\mathbb{Z}$-basis of $\mathbb{Z}^n$. However, the upper bound on the corresponding denominators $q_1, \ldots, q_n$ is necessarily larger than Q^{n-1}, and is given by a function of Q that we can call here Ψ'_{ω_0} (see [BF13] for more precise and general information, but note that in this reference, Ψ'_{ω_0} was denoted by Ψ_{ω_0} and Ψ_{ω_0}, which we defined in (11), was denoted by Ψ'_{ω_0}). A consequence of the main Diophantine result of [BF13] is that this function Ψ'_{ω_0} is in fact essentially equivalent to the function Ψ_{ω_0}.

PROPOSITION 21. *Let $\omega_0 = (1, \bar\omega_0) \in \mathbb{R}^n$ be a non-resonant vector, with $\bar\omega_0 \in [-1,1]^{n-1}$. For any $Q \geq n+2$, there exist n rational vectors $v_1, \ldots, v_n$, of denominators $q_1, \ldots, q_n$, such that $q_1 v_1, \ldots, q_n v_n$ form a $\mathbb{Z}$-basis of $\mathbb{Z}^n$ and for $j \in \{1, \ldots, n\}$,*

$$|\omega_0 - v_j| \lessdot (q_j Q)^{-1}, \quad 1 \leq q_j \lessdot \Psi(Q).$$

For a proof of the above proposition with Ψ_{ω_0} instead of Ψ, we refer to [**BF13**], Theorem 2.1 and Proposition 2.3; now by (12), $\Psi_{\omega_0} \leq \Psi$ and so one may replace Ψ_{ω_0} by Ψ.

Now given a q-rational vector v and a smooth function H defined on $\mathbb{T}^n \times D_{r,h}$, we define

$$(46) \qquad [H]_v(\theta, I, \omega) = \int_0^1 H(\theta + tqv, I, \omega) dt.$$

Given n rational vectors $v_1, \ldots, v_n$, we let $[H]_{v_1, \ldots, v_d} = [\cdots [H]_{v_1} \cdots]_{v_d}$. Finally we define

$$(47) \qquad [H](I, \omega) = \int_{\mathbb{T}^n} H(\theta, I, \omega) d\theta.$$

The following proposition is a consequence of the fact that the vectors $q_1 v_1, \ldots, q_n v_n$ form a $\mathbb{Z}$-basis of $\mathbb{Z}^n$.

PROPOSITION 22 ([**Bou13b**, Corollary 6]). *Let $v_1, \ldots, v_n$ be rational vectors, of denominators $q_1, \ldots, q_n$, such that $q_1 v_1, \ldots, q_n v_n$ form a $\mathbb{Z}$-basis of $\mathbb{Z}^n$, and H a function defined on $\mathbb{T}^n \times D_{r,h}$. Then*

$$[H]_{v_1, \ldots, v_n} = [H].$$

3. KAM step

Now we describe an elementary step of our iterative procedure. Such a step consists of pulling back the Hamiltonian H by a transformation of the form

$$\mathcal{F} = (\Phi, \varphi) : (\theta, I, \omega) \mapsto (\Phi(\theta, I, \omega), \varphi(\omega));$$

Φ is a parameter-depending change of coordinates and φ a change of parameters. Moreover, our change of coordinates will be of the form

$$\Phi(\theta, I, \omega) = \Phi_\omega(\theta, I) = (\theta + E(\theta, \omega), I + F(\theta, \omega) \cdot I + G(\theta, \omega))$$

with

$$E : \mathbb{T}^n \times D_h \to \mathbb{R}^n, \quad F : \mathbb{T}^n \times D_h \to M_n(\mathbb{R}), \quad G : \mathbb{T}^n \times D_h \to \mathbb{R}^n$$

and for each fixed parameter ω, Φ_ω will be symplectic. For simplicity, we shall write $\Phi = (E, F, G)$; the composition of such transformations $\mathcal{F} = (\Phi, \varphi) = (E, F, G, \varphi)$ is again a transformation of the same form, and we shall denote by $\mathcal{G}$ the groupoid of such transformations.

PROPOSITION 23. *Let H be as in $(**)$, with $M = (M_l)_{l \in \mathbb{N}}$ satisfying (H1) and (H2) and $\omega_0 = (1, \bar\omega_0) \in \mathbb{R}^n$ non-resonant. Consider $0 < \sigma < 1$, $0 < \delta < r$, $Q \geq n+2$, and assume that*

$$(48) \quad \sqrt\varepsilon \leq \mu \leq h/2, \quad \sqrt\varepsilon \leq r, \quad h \lessdot (Q\Psi(Q))^{-1}, \quad r\mu \lessdot \delta(Q\Psi(Q))^{-1}, \quad (1+\eta) \lessdot QsC(\sigma)^{-1}.$$

Then there exists an $(M, s(1-\sigma)^{2n+1})$-ultra-differentiable transformation

$$\mathcal{F} = (\Phi, \varphi) = (E, F, G, \varphi) : \mathbb{T}^n \times D_{r-\delta, h/2} \to \mathbb{T}^n \times D_{r,h} \in \mathcal{G},$$

with the estimates

$$(49) \quad \begin{cases} |E|_{s(1-\sigma)^{2n}} \lessdot \Psi(Q)\mu, & |\nabla E|_{s(1-\sigma)^{2n+1}} \lessdot s^{-1}C(\sigma)\Psi(Q)\mu, \\ |F|_{s(1-\sigma)^{2n}} \lessdot s^{-1}C(\sigma)\Psi(Q)\mu, & |\nabla F|_{s(1-\sigma)^{2n+1}} \lessdot s^{-2}C(\sigma)^2\Psi(Q)\mu, \\ |G|_{s(1-\sigma)^{2n}} \lessdot s^{-1}C(\sigma)\Psi(Q)\varepsilon, & |\nabla G|_{s(1-\sigma)^{2n+1}} \lessdot s^{-2}C(\sigma)^2\Psi(Q)\varepsilon, \\ |\varphi - \mathrm{Id}|_{s(1-\sigma)^{2n}} \leq \mu, & |\nabla\varphi - \mathrm{Id}|_{s(1-\sigma)^{2n+1}} \lessdot s^{-1}C(\sigma)\mu \end{cases}$$

such that

$$H \circ \mathcal{F}(\theta, I, \omega) = \underbrace{e^+(\omega) + \omega \cdot I}_{N^+(I,\omega)} + \underbrace{A^+(\theta) + B^+(\theta) \cdot I}_{P^+(\theta,I,\omega)} + \underbrace{M^+(\theta, I, \omega) \cdot I^2}_{R^+(\theta,I,\omega)},$$

with the estimates

(50)
$$
\begin{cases}
|A^+|_{s(1-\sigma)^{2n+1}} \leq \varepsilon/16, \quad |B^+|_{s(1-\sigma)^{2n+1}} \leq \mu/4, \\
|e^+ - e \circ \varphi|_{s(1-\sigma)^{2n+1}} \leq |A|_s, \quad |\nabla_I^2 R^+ - \nabla_I^2 R \circ \mathcal{F}|_{s(1-\sigma)^{2n+1}} \lessdot \eta |F|_{s(1-\sigma)^{2n}}.
\end{cases}
$$

PROOF. We divide the proof of the KAM step into five small steps. Except for the last one, the parameter $\omega \in D_h$ will be fixed, so for simplicity, in the first four steps we will drop the dependence on the parameter $\omega \in D_h$. Let us first notice that (48) clearly implies the following seven inequalities:

(51) $$h \lessdot (Q\Psi(Q))^{-1}$$

(52) $$\Psi(Q)\mu \lessdot sC(\sigma)^{-1}$$

(53) $$\varepsilon \leq r\mu$$

(54) $$r\mu s^{-1}C(\sigma)\Psi(Q) \lessdot \delta$$

(55) $$\mu \lessdot (Q\Psi(Q))^{-1}$$

(56) $$(1+\eta) \lessdot QsC(\sigma)^{-1}$$

(57) $$\mu \leq h/2.$$

It is also important to notice that the implicit constant appearing in (56) is independent of the other ones; we may choose it as large as we want without affecting the other implicit constants. In the first three steps, the term R which contains terms of order at least 2 in I will be ignored, that is we will only consider $\hat{H} = H - R = N + P$.

1. Rational approximations of ω_0 and $\omega \in D_h$

Since ω_0 is non-resonant, given $Q \geq n+2$, we can apply Proposition 21: there exist n rational vectors $v_1, \ldots, v_n$, of denominators $q_1, \ldots, q_n$, such that $q_1 v_1, \ldots, q_n v_n$ form a $\mathbb{Z}$-basis of $\mathbb{Z}^n$ and for $j \in \{1, \ldots, n\}$,
$$|\omega_0 - v_j| \lessdot (q_j Q)^{-1}, \quad 1 \leq q_j \lessdot \Psi(Q).$$
For any $\omega \in D_h$, using (51) and $q_j \lessdot \Psi(Q)$, we have

(58) $$|\omega - v_j| \leq |\omega - \omega_0| + |\omega_0 - v_j| \lessdot h + (q_j Q)^{-1} \lessdot (Q\Psi(Q))^{-1} + (q_j Q)^{-1} \lessdot (q_j Q)^{-1}.$$

2. Successive rational averagings

Let us set $A_1 := A$, $B_1 := B$ so that $P_1(\theta, I) := A_1(\theta) + B_1(\theta) \cdot I$ satisfies $P_1 = P$. Recalling that $[\,.\,]_v$ denotes the averaging along the periodic flow associated to a periodic vector $v \in \mathbb{R}^n$ (see (46)), we define inductively, for $1 \leq j \leq n$,
$$A_{j+1} := [A_j]_{v_j}, \quad B_{j+1} := [B_j]_{v_j}, \quad P_{j+1} := [P_j]_{v_j}$$
so in particular $P_j(\theta, I) = A_j(\theta) + B_j(\theta) \cdot I$ for $1 \leq j \leq n$. Let us also define X_j, for $1 \leq j \leq n$, by
$$X_j(\theta, I) := C_j(\theta) + D_j(\theta) \cdot I$$
where
$$C_j(\theta) = q_j \int_0^1 (A_j - A_{j+1})(\theta + t q_j v_j) t \, dt, \quad D_j(\theta) = q_j \int_0^1 (B_j - B_{j+1})(\theta + t q_j v_j) t \, dt.$$
If we further define N_j by $N_j(I) = e(\omega) + v_j \cdot I$, it is then easy to check, by a simple integration by parts, that the equations

(59) $$\{C_j, N_j\} = A_j - A_{j+1}, \quad \{D_j, N_j\} = B_j - B_{j+1}, \quad 1 \leq j \leq n,$$

are satisfied and then

(60) $$\{X_j, N_j\} = P_j - P_{j+1}, \quad 1 \leq j \leq n,$$

are also satisfied, where $\{\,.\,,\,.\,\}$ denotes the usual Poisson bracket. Moreover, we have the estimates

(61) $$|A_j|_s \leq |A|_s \leq \varepsilon, \quad |B_j|_s \leq |B|_s \leq \mu,$$

and then

(62) $$\quad\quad\quad\quad |C_j|_s \le q_j |A_j|_s \le q_j \varepsilon \lessdot \Psi(Q)\varepsilon, \quad |D_j|_s \le q_j |B_j|_s \le q_j \mu \lessdot \Psi(Q)\mu.$$

Next, for any $0 \le j \le n$, define

$$r_j := r - n^{-1}j\delta, \quad \hat{s}_j := s(1-\sigma)^{2j}, \quad s_j = s(1-\sigma)^{2j+1}.$$

Let X_j^t be the time-t map of the Hamiltonian flow of X_j. Using (62), together with inequalities (52), (53), (54) and the fact that $C(\sigma)^{-1} \lessdot \sigma$ (this is (6)), the condition (34) and (36) of Proposition 18, § 1, are satisfied, so the latter proposition can be applied: for $1 \le j \le n$, X_j^t is $(M, \hat{s}_j)$-ultra-differentiable, it maps $\mathbb{T}^n \times D_{r_j}$ into $\mathbb{T}^n \times D_{r_{j-1}}$ for all $t \in [0,1]$ and it is of the form

$$X_j^t(\theta, I) = (\theta + E_j^t(\theta), I + F_j^t(\theta) \cdot I + G_j^t(\theta))$$

with

(63) $$\begin{cases} |E_j^t|_{\hat{s}_j} \le |D_j|_{s_{j-1}} \lessdot \Psi(Q)\mu, \quad |\nabla E_j^t|_{s_j} \lessdot s^{-1}C(\sigma)\Psi(Q)\mu \\ |F_j^t|_{\hat{s}_j} \lessdot s^{-1}C(\sigma)|D_j|_{s_{j-1}} \lessdot s^{-1}C(\sigma)\Psi(Q)\mu, \quad |\nabla F_j^t|_{s_j} \lessdot s^{-2}C(\sigma)^2\Psi(Q)\mu \\ |G_j^t|_{\hat{s}_j} \lessdot s^{-1}C(\sigma)|C_j|_{s_{j-1}} \lessdot s^{-1}C(\sigma)\Psi(Q)\varepsilon, \quad |\nabla G_j^t|_{s_j} \lessdot s^{-2}C(\sigma)^2\Psi(Q)\varepsilon. \end{cases}$$

Now we define $\Phi^0 := \mathrm{Id}$ to be the identity and inductively $\Phi^j := \Phi^{j-1} \circ X_j^1$ for $1 \le j \le n$. Then Φ^j maps $\mathbb{T}^n \times D_{r_j}$ into $\mathbb{T}^n \times D_r$ and one easily checks, by induction using the estimates (63), that Φ^j is still of the form

$$\Phi^j(\theta, I) = (\theta + E^j(\theta), I + F^j(\theta) \cdot I + G^j(\theta))$$

with the estimates, for $j = 1, ..., n$,

(64) $$\begin{cases} |E^j|_{\hat{s}_j} \lessdot \Psi(Q)\mu, \quad |\nabla E^j|_{s_j} \lessdot s^{-1}C(\sigma)\Psi(Q)\mu, \\ |F^j|_{\hat{s}_j} \lessdot s^{-1}C(\sigma)\Psi(Q)\mu, \quad |\nabla F^j|_{s_j} \lessdot s^{-2}C(\sigma)^2\Psi(Q)\mu, \\ |G^j|_{\hat{s}_j} \lessdot s^{-1}C(\sigma)\Psi(Q)\varepsilon, \quad |\nabla G^j|_{s_j} \lessdot s^{-2}C(\sigma)^2\Psi(Q)\varepsilon. \end{cases}$$

3. New Hamiltonian

Let us come back to the Hamiltonian $\hat{H} = H - R = N + P = N + P_1$. We claim that for all $0 \le j \le n$, we have

$$\hat{H} \circ \Phi^j = N + P_{j+1} + P_{j+1}^+, \quad P_{j+1}^+(\theta, I) = A_{j+1}^+(\theta) + B_{j+1}^+(\theta) \cdot I$$

with the estimates

(65) $$\quad\quad\quad\quad |A_{j+1}^+|_{\hat{s}_j} \lessdot (Qs)^{-1}C(\sigma)\varepsilon, \quad |B_{j+1}^+|_{\hat{s}_j} \lessdot (Qs)^{-1}C(\sigma)\mu.$$

Let us prove the claim by induction on $0 \le j \le n$. For $j = 0$, we may set $P_1^+ := 0$ and there is nothing to prove. So let us assume that the claim is true for some $j - 1 \ge 0$, and we need to show it is still true for $j \ge 1$. By this inductive assumption, we have

$$\hat{H} \circ \Phi^j = \hat{H} \circ \Phi^{j-1} \circ X_j^1 = (N + P_j + P_j^+) \circ X_j^1$$

with

(66) $$\quad\quad\quad\quad |A_j^+|_{\hat{s}_{j-1}} \lessdot (Qs)^{-1}C(\sigma)\varepsilon, \quad |B_j^+|_{\hat{s}_{j-1}} \lessdot (Qs)^{-1}C(\sigma)\mu.$$

Let $S_j = \omega \cdot I - v_j \cdot I$ so that $N = N_j + S_j$ and thus

$$\hat{H} \circ \Phi^j = (N_j + S_j + P_j + P_j^+) \circ X_j^1 = (N_j + S_j + P_j) \circ X_j^1 + P_j^+ \circ X_j^1.$$

Let us consider the first summand of the last sum. Using the equality (60), a standard computation based on Taylor's formula with integral remainder gives

$$(N_j + S_j + P_j) \circ X_j^1 = N + [P_j]_{v_j} + \tilde{P}_{j+1} = N + P_{j+1} + \tilde{P}_{j+1}$$

with

$$\tilde{P}_{j+1} = \int_0^1 W_{j+1}^t \circ X_j^t dt, \quad W_{j+1}^t := \{(1-t)P_{j+1} + tP_j + S_j, X_j\}.$$

Clearly, W_{j+1}^t is still of the form

$$W_{j+1}^t(\theta, I) = W_{j+1}^t(\theta, 0) + \nabla_I W_{j+1}^t(\theta, 0) \cdot I$$

as this is true for P_j, S_j, X_j and that this form is preserved under Poisson bracket. Using the estimates for $P_j(\theta, 0)$, $\nabla_I P_j(\theta, 0)$, $X_j(\theta, 0)$, $\nabla_I X_j(\theta, 0)$ (given respectively in (61) and in (62)), the fact that

$$S_j(\theta, 0) = 0, \quad \nabla_I S_j(\theta, 0) = \omega - v_j$$

with the inequality (58), and the estimates for the derivatives and the product of Gevrey functions (given respectively in Proposition 8, Corollary 9 and Proposition 10, Corollary 13, § 1), one finds, for all $t \in [0, 1]$

$$
\begin{aligned}
|W_{j+1}^t(\theta, 0)|_{\hat{s}_j} &< (s^{-1}C(\sigma)q_j\varepsilon\mu + s^{-1}C(\sigma)q_j\varepsilon\mu + s^{-1}C(\sigma)q_j\varepsilon(q_jQ)^{-1}) \\
&< s^{-1}C(\sigma)q_j\varepsilon\mu + (Qs)^{-1}C(\sigma)\varepsilon.
\end{aligned}
$$

Since $q_j < \Psi(Q)$, using (55) the latter estimate reduces to

$$|W_{j+1}^t(\theta, 0)|_{\hat{s}_j} < (Qs)^{-1}C(\sigma)\varepsilon.$$

Similarly, one obtains

$$|\nabla_I W_{j+1}^t(\theta, 0)|_{\hat{s}_j} < (Qs)^{-1}C(\sigma)\mu.$$

Then, using the expression of X_j^t and the associated estimates (63), a direct computation, still using (55), gives

$$|\tilde{P}_{j+1}(\theta, 0)|_{\hat{s}_j} < (Qs)^{-1}C(\sigma)\varepsilon$$

and

$$|\nabla_I \tilde{P}_{j+1}(\theta, 0)|_{\hat{s}_j} < (Qs)^{-1}C(\sigma)\mu.$$

Using again the estimates of X_j^t given by (63), and the inductive assumption (66), we also find

$$|P_j^+ \circ X_j^1(\theta, 0)|_{\hat{s}_j} < (Qs)^{-1}C(\sigma)\varepsilon$$

and

$$|\nabla_I(P_j^+ \circ X_j^1)(\theta, 0)|_{\hat{s}_j} < (Qs)^{-1}C(\sigma)\mu.$$

Eventually, we may define

$$P_{j+1}^+ := \tilde{P}_{j+1} + P_j^+ \circ X_j^1$$

so that

$$\hat{H} \circ \Phi^j = N + P_{j+1} + P_{j+1}^+, \quad P_{j+1}^+(\theta, I) = A_{j+1}^+(\theta) + B_{j+1}^+(\theta) \cdot I$$

and these last estimates imply that

$$|A_{j+1}^+|_{\hat{s}_j} < (Qs)^{-1}C(\sigma)\varepsilon, \quad |B_{j+1}^+|_{\hat{s}_j} < (Qs)^{-1}C(\sigma)\mu.$$

The claim is proved. So we may set

$$\Phi := \Phi^n, \quad (E, F, G) := (E^n, F^n, G^n),$$

with, as (64) tells us with $j = n$,

$$
(67) \quad
\begin{cases}
|E|_{\hat{s}_n} < \Psi(Q)\mu, \quad |\nabla E|_{s_n} < s^{-1}C(\sigma)\Psi(Q)\mu \\
|F|_{\hat{s}_n} < s^{-1}C(\sigma)\Psi(Q)\mu, \quad |\nabla F|_{s_n} < s^{-2}C(\sigma)^2\Psi(Q)\mu \\
|G|_{\hat{s}_n} < s^{-1}C(\sigma)\Psi(Q)\varepsilon, \quad |\nabla G|_{s_n} < s^{-2}C(\sigma)^2\Psi(Q)\varepsilon.
\end{cases}
$$

Observe that $P_{n+1} = [\cdots[P]_{v_1}\cdots]_{v_n} = [P]_{v_1,\ldots,v_n}$, and thus by Proposition 22, $P_{n+1} = [P]$, and as a consequence

$$\hat{H} \circ \Phi(\theta, I) = e + \omega \cdot I + [A] + [B] \cdot I + A_{n+1}^+(\theta) + B_{n+1}^+(\theta) \cdot I$$

with the estimates

$$(68) \quad |A_{n+1}^+|_{\hat{s}_n} < (Qs)^{-1}C(\sigma)\varepsilon, \quad |B_{n+1}^+|_{\hat{s}_n} < (Qs)^{-1}C(\sigma)\mu.$$

4. Estimate of the remainder

Now we take into account the remainder term R that we previously ignored: we have $H = \hat{H} + R$, and therefore

$$H \circ \Phi(\theta, I) = e + \omega \cdot I + [A] + [B] \cdot I + A_{n+1}^+(\theta) + B_{n+1}^+(\theta) \cdot I + R \circ \Phi(\theta, I).$$

Let us decompose

$$R \circ \Phi(\theta, I) = \underbrace{R \circ \Phi(\theta, 0)}_{R_0(\theta)} + \underbrace{\nabla_I (R \circ \Phi)(\theta, 0)}_{R_1(\theta)} \cdot I + \tilde{R}(\theta, I)$$

and let us define

$$\tilde{A} := A_{n+1}^+ + R_0, \quad \tilde{B} := B_{n+1}^+ + R_1.$$

We have $R(\theta, I) = M(\theta, I) \cdot I^2$ and as H and R differ only by terms of order at most one in I, $\nabla_I^2 H = \nabla_I^2 R$ so

$$M(\theta, I) = \int_0^1 (1-t) \nabla_I^2 H(\theta, tI) dt = \int_0^1 (1-t) \nabla_I^2 R(\theta, tI) dt$$

and therefore $|M|_s \leq \eta$. Then, as $\Phi(\theta, 0) = (\theta + E(\theta), G(\theta))$, we have the expression

$$R_0(\theta) = R(\Phi(\theta, 0)) = M(\theta + E(\theta), G(\theta)) \cdot G(\theta)^2$$

and so using the above estimate on M, together with the estimates on E, G and the estimates for the product and composition of Gevrey functions (given respectively in Proposition 10 and Proposition 16, § 1), we find

$$|R_0|_{\hat{s}_n} \lessdot \eta |G|_{\hat{s}_n}^2 \lessdot \eta (s^{-1} C(\sigma) \Psi(Q) \mu)^2 \varepsilon \lessdot \eta (Qs)^{-2} C(\sigma)^2 \varepsilon \lessdot \eta (Qs)^{-1} C(\sigma) \varepsilon.$$

Then, we have $\nabla_I R(\theta, I) = \hat{M}(\theta, I) \cdot I^2$ with

$$\hat{M}(\theta, I) = \int_0^1 \nabla_I^2 H(\theta, tI) dt = \int_0^1 \nabla_I^2 R(\theta, tI) dt$$

and hence $|\hat{M}|_s \leq \eta$ also. Since

$$(69) \qquad |\nabla_I \Phi - \mathrm{Id}|_{\hat{s}_n} = |F|_{\hat{s}_n} \lessdot s^{-1} C(\sigma) \Psi(Q) \mu \lessdot 1$$

we obtain, using the fact that $\varepsilon \leq \mu^2$ and proceeding as before,

$$|R_1|_{\hat{s}_n} \lessdot \eta |G|_{\hat{s}_n} \lessdot \eta s^{-1} C(\sigma) \Psi(Q) \varepsilon \lessdot \eta (s^{-1} C(\sigma) \Psi(Q) \mu) \mu \lessdot \eta (Qs)^{-1} C(\sigma) \mu.$$

These last estimates on R_0 and R_1, together with (68), imply

$$|\tilde{A}|_{\hat{s}_n} \lessdot (1+\eta)(Qs)^{-1} C(\sigma) \varepsilon, \quad |\tilde{B}|_{\hat{s}_n} \lessdot (1+\eta)(Qs)^{-1} C(\sigma) \mu.$$

We can finally now use (56) to ensure that

$$(70) \qquad |\tilde{A}|_{\hat{s}_n} \leq \varepsilon/16, \quad |\tilde{B}|_{\hat{s}_n} \leq \mu/4.$$

It is important to recall here that we may choose the implicit constant in (56) as large as we want (in order to achieve (70)) without affecting any of the other implicit constants. Then observe also that $H \circ \Phi$ and $\tilde{R}$ differ only by terms of order at most one in I, so

$$\nabla_I^2 (H \circ \Phi) = \nabla_I^2 \tilde{R}, \quad \nabla_I^2 H = \nabla_I^2 R$$

and therefore using the formula for the Hessian of a composition, (69) and the fact that $\nabla_I^2 \Phi$ is identically zero, one finds

$$(71) \qquad |\nabla_I^2 \tilde{R} - \nabla_I^2 R \circ \Phi|_{\hat{s}_n} \lessdot \eta |F|_s.$$

We also set $\tilde{e} := e + [A]$ and observe that

$$(72) \qquad |\tilde{e} - e|_{\hat{s}_n} \leq |[A]|_s \leq |A|_s.$$

5. Change of frequencies and final estimates

Let us now write explicitly the dependence on the parameter $\omega \in D_h$: we have

$$H \circ \Phi(\theta, I, \omega) = \tilde{e}(\omega) + (\omega + [B](\omega)) \cdot I + \tilde{A}(\theta, \omega) + \tilde{B}(\theta, \omega) \cdot I + \tilde{R}(\theta, I, \omega).$$

Consider the map $\phi(\omega) := \omega + [B](\omega)$, it satisfies

$$|\phi - \mathrm{Id}|_s \leq |[B]|_s \leq |B|_s \leq \mu$$

and therefore the conditions (41) of Proposition 20 are satisfied: the first condition of (41) follows, from instance, from condition (52) and the fact that $\Psi(Q) \geq Q \geq 1$, whereas the second condition

of (41) is implied by condition (57). Hence Proposition 20 applies and one finds a unique $\varphi \in G_{\hat{s}_n}(D_{h/2}, D_h)$ such that $\phi \circ \varphi = \mathrm{Id}$ and

$$(73) \qquad\qquad |\varphi - \mathrm{Id}|_{\hat{s}_n} \leq |\phi - \mathrm{Id}|_s \leq \mu.$$

We do have $\varphi(\omega) + [B(\varphi(\omega))] = \omega$ and thus, setting $\mathcal{F} := (\Phi, \varphi)$, this implies that

$$H \circ \mathcal{F}(\theta, I, \omega) = H \circ \Phi(\theta, I, \varphi(\omega)) = \tilde{e}(\varphi(\omega)) + \omega \cdot I + \tilde{A}(\theta, \varphi(\omega)) + \tilde{B}(\theta, \varphi(\omega)) \cdot I + \tilde{R}(\theta, I, \varphi(\omega))$$

and in the end we set

$$e^+ := \tilde{e} \circ \varphi, \quad A^+ := \tilde{A} \circ \varphi, \quad B^+ := \tilde{B} \circ \varphi, \quad R^+ := \tilde{R} \circ \varphi.$$

Using once again Proposition 16, the inequalities (70), (71) and (72) imply

$$\begin{cases} |A^+|_{s_n} = |\tilde{A} \circ \varphi|_{s_n} \leq |\tilde{A}|_{\hat{s}_n} \leq \varepsilon/16, \\ |B^+|_{s_n} = |\tilde{B} \circ \varphi|_{s_n} \leq |\tilde{B}|_{\hat{s}_n} \leq \mu/4, \\ |e^+ - e \circ \varphi|_{s_n} = |(\tilde{e} - e) \circ \varphi|_{s_n} \leq |\tilde{e} - e|_{\hat{s}_n} \leq |A|_s, \\ |\nabla_I^2 R^+ - \nabla_I^2 R \circ \mathcal{F}|_{s_n} = |(\nabla_I^2 \tilde{R} - \nabla_I^2 R \circ \Phi) \circ \varphi|_{s_n} \leq |\nabla_I^2 \tilde{R} - \nabla_I^2 R \circ \Phi|_{\hat{s}_n} \lessdot \eta |F|_{\hat{s}_n}, \end{cases}$$

which are the estimates (50) we needed to prove. The transformation $\mathcal{F} = (\Phi, \varphi) = (E, F, G, \varphi) \in \mathcal{G}$ maps $\mathbb{T}^n \times D_{r-\delta,h/2}$ into $\mathbb{T}^n \times D_{r,h}$ and it follows from (67) and (73) that

$$\begin{cases} |E|_{\hat{s}_n} \lessdot \Psi(Q)\mu, \quad |\nabla E|_{s_n} \lessdot s^{-1}C(\sigma)\Psi(Q)\mu, \\ |F|_{\hat{s}_n} \lessdot s^{-1}C(\sigma)\Psi(Q)\mu, \quad |\nabla F|_{s_n} \lessdot s^{-2}C(\sigma)^2\Psi(Q)\mu, \\ |G|_{\hat{s}_n} \lessdot \sigma^{-\alpha}\Psi(Q)\varepsilon, \quad |\nabla G|_{s_n} \lessdot s^{-2}C(\sigma)^2\Psi(Q)\varepsilon, \\ |\varphi - \mathrm{Id}|_{\hat{s}_n} \leq \mu, \quad |\nabla \varphi - \mathrm{Id}|_{s_n} \lessdot s^{-1}C(\sigma)\mu \end{cases}$$

which were the wanted estimates (49). This concludes the proof. $\qquad\qquad\square$

4. Iterations and convergence

We now define, for $i \in \mathbb{N}$, the following decreasing geometric sequences:

$$(74) \qquad\qquad \varepsilon_i := 16^{-i}\varepsilon, \quad \mu_i := 4^{-i}\mu, \quad \delta_i := 2^{-i-2}r, \quad h_i = 2^{-i}h.$$

Next, for a constant Q_0 to be chosen below, we define Δ_i and Q_i, $i \in \mathbb{N}$, by

$$(75) \qquad\qquad \Delta_i = 2^i \Delta(Q_0), \quad Q_i = \Delta^{-1}(\Delta_i) = \Delta^{-1}(2^i \Delta(Q_0)).$$

Let us now choose $0 < c_2 \leq 1$, which depends only on n, for which the last inequality of (48) is satisfied with implicit constant $2c_2$, that is

$$(76) \qquad\qquad (1+\eta)(2c_2)^{-1} \leq QsC(\sigma)^{-1}.$$

We can now define σ_i, $i \in \mathbb{N}$, by

$$(77) \qquad\qquad \sigma_i = C^{-1}(c_2(1+\eta)^{-1}sQ_i)$$

and finally, we define $\hat{s}_i$, s_i and r_i, $i \in \mathbb{N}$, by

$$(78) \quad \hat{s}_0 = s_0 = s, \quad \hat{s}_{i+1} = \hat{s}_i(1-\sigma_i)^{2n}, \quad s_{i+1} = s_i(1-\sigma_i)^{2n+1}, \quad r_0 = r, \quad r_{i+1} = r_i - \delta_i.$$

Obviously, we have

$$\lim_{i \to +\infty} r_i = r - \sum_{i \in \mathbb{N}} \delta_i = r/2.$$

We claim that, assuming that ω_0 satisfies $(\mathrm{BR_M})$, if we choose Q_0 as in (44) we have

$$(79) \qquad\qquad \lim_{i \to +\infty} s_i \geq s/2 \iff \prod_{i \in \mathbb{N}} (1-\sigma_i)^{2n+1} \geq 1/2.$$

Indeed, we have

$$
\begin{aligned}
\sum_{i\in\mathbb{N}} \sigma_i &= \sigma_0 + \sum_{i\geq 1} \sigma_i \\
&= C^{-1}(c_2(1+\eta)^{-1}sQ_0) + \sum_{i\geq 1} C^{-1}(c_2(1+\eta)^{-1}s\Delta^{-1}(2^i\Delta(Q_0))) \\
&\leq C^{-1}(c_2(1+\eta)^{-1}sQ_0) + \int_0^{+\infty} C^{-1}(c_2(1+\eta)^{-1}s\Delta^{-1}(2^y\Delta(Q_0)))dy \\
&= C^{-1}(c_2(1+\eta)^{-1}sQ_0) + (\ln 2)^{-1}\int_{\Delta(Q_0)}^{+\infty} C^{-1}(c_2(1+\eta)^{-1}s\Delta^{-1}(x))\frac{dx}{x} \\
&\leq \ln 2(4n+2)^{-1}
\end{aligned}
$$

where we used (44) in the last line. In particular, for any $i \in \mathbb{N}$ we have $\sigma_i \leq \ln 2(4n+2)^{-1}$ which implies

$$
\ln(1-\sigma_i)^{2n+1} = (2n+1)\ln(1-\sigma_i) \geq -(4n+2)\sigma_i
$$

and thus

$$
\sum_{i\in\mathbb{N}} \ln(1-\sigma_i)^{2n+1} \geq -(4n+2)\sum_{i\in\mathbb{N}} \sigma_i \geq -\ln 2
$$

and finally

$$
\prod_{i\in\mathbb{N}}(1-\sigma_i)^{2n+1} = \exp\left(\sum_{i\in\mathbb{N}}\ln(1-\sigma_i)^{2n+1}\right) \geq 1/2.
$$

This shows that (79) holds true with our choice of σ_i. Applying inductively Proposition 23 we will easily obtain the following proposition.

PROPOSITION 24. *Let H be as in $(**)$, with $M = (M_l)_{l\in\mathbb{N}}$ satisfying (H1) and (H2) and ω_0 satisfying $(\mathrm{BR_M})$. Fix $Q_0 \geq n+2$ so that (44) is satisfied, and assume that*

$$
(80) \qquad \sqrt{\varepsilon} \leq \mu \leq h/2, \quad \sqrt{\varepsilon} \leq r, \quad h \lessdot \Delta(Q_0)^{-1}.
$$

Then, for each $i \in \mathbb{N}$, there exists an (M, s_i)-ultra-differentiable transformation

$$
\mathcal{F}^i = (\Phi^i, \varphi^i) = (E^i, F^i, G^i, \varphi^i) : \mathbb{T}^n \times D_{r_i,h_i} \to \mathbb{T}^n \times D_{r,h} \in \mathcal{G},
$$

such that $\mathcal{F}^{i+1} = \mathcal{F}^i \circ \mathcal{F}_{i+1}$, with

$$
\mathcal{F}_{i+1} = (\Phi_{i+1}, \varphi_{i+1}) = (E_{i+1}, F_{i+1}, G_{i+1}, \varphi_{i+1}) : \mathbb{T}^n \times D_{r_{i+1},h_{i+1}} \to \mathbb{T}^n \times D_{r_i,h_i} \in \mathcal{G},
$$

satisfying the following estimates

$$
(81) \qquad \begin{cases}
|E_{i+1}|_{\hat{s}_{i+1}} \lessdot \Psi(Q_i)\mu_i, \quad |\nabla E_{i+1}|_{s_{i+1}} \lessdot s_i^{-1}C(\sigma_i)\Psi(Q_i)\mu_i, \\
|F_{i+1}|_{\hat{s}_{i+1}} \lessdot s_i^{-1}C(\sigma_i)\Psi(Q_i)\mu_i, \quad |\nabla F_{i+1}|_{s_{i+1}} \lessdot s_i^{-2}C(\sigma_i)^2\Psi(Q_i)\mu_i, \\
|G_{i+1}|_{\hat{s}_{i+1}} \lessdot s_i^{-1}C(\sigma_i)\Psi(Q_i)\varepsilon_i, \quad |\nabla G_{i+1}|_{s_{i+1}} \lessdot s_i^{-2}C(\sigma_i)^2\Psi(Q_i)\varepsilon_i, \\
|\varphi_{i+1} - \mathrm{Id}|_{\hat{s}_{i+1}} \leq \mu_i, \quad |\nabla\varphi_{i+1} - \mathrm{Id}|_{s_{i+1}} \lessdot s_i^{-1}C(\sigma_i)\mu_i
\end{cases}
$$

and such that

$$
H \circ \mathcal{F}^i(\theta, I, \omega) = \underbrace{e^i(\omega) + \omega \cdot I}_{N^i(I,\omega)} + \underbrace{A^i(\theta) + B^i(\theta)\cdot I}_{P^i(\theta,I,\omega)} + \underbrace{M^i(\theta, I, \omega)\cdot I^2}_{R^i(\theta,I,\omega)}
$$

with the estimates

$$
(82) \qquad \begin{cases}
|A^i|_{s_{i+1}} \leq \varepsilon_i, \quad |B^i|_{s_{i+1}} \leq \mu_i, \\
|e^{i+1} - e^i \circ \varphi_{i+1}|_{s_{i+1}} \leq |A^i|_{\alpha,s_i}, \\
|\nabla_I^2 R^{i+1} - \nabla_I^2 R^i \circ \mathcal{F}_{i+1}|_{s_{i+1}} \lessdot \eta|F_{i+1}|_{\hat{s}_{i+1}}.
\end{cases}
$$

Let us emphasize that the implicit constants in the above proposition depend only on n and are thus independent of $i \in \mathbb{N}$.

PROOF. For $i = 0$, we let $\mathcal{F}^0$ be the identity, $e^0 := e$, $A^0 := A$, $B^0 := B$, $R^0 := R$, $M^0 := M$ and there is nothing to prove. The general case follows by an easy induction. Indeed, assume that the statement holds true for some $i \in \mathbb{N}$ so that $H \circ \mathcal{F}^i$ is (M, s_i)-ultra-differentiable on the domain $\mathbb{T}^n \times D_{r_i, s_i}$. We want to apply Proposition 23 to this Hamiltonian, with $\varepsilon = \varepsilon_i$, $\mu = \mu_i$, $r = r_i$, $s = s_i$, $h = h_i$, $\sigma = \sigma_i$ and $Q = Q_i$. First, by our choice of Q_0 and δ_0 it is clear that $0 < \sigma_i < 1$, $0 < \delta_i < r_i$, and $Q_i \geq n + 2$. Then, recalling the definition of the constant c_2, we need to check that the conditions

$$\begin{cases} \sqrt{\varepsilon_i} \leq \mu_i \leq h_i/2, & \sqrt{\varepsilon_i} \leq r_i, & h_i \cdot< \Delta(Q_i)^{-1}, \\ r_i \mu_i \cdot< \delta_i \Delta(Q_i)^{-1}, & (1+\eta)(2c_2)^{-1} \leq Q_i s_i C(\sigma_i)^{-1} \end{cases}$$

are satisfied. From the definition of Q_i and σ_i we have

$$\Delta(Q_i) = \Delta(\Delta^{-1}(\Delta_i)) = \Delta_i, \quad C(\sigma_i) = C(C^{-1}(c_2(1+\eta)^{-1} s Q_i)) = c_2(1+\eta)^{-1} s Q_i$$

and as $2^{-i} r \leq r_i \leq r$, it is sufficient to check the conditions

(83) $$\sqrt{\varepsilon_i} \leq \mu_i \leq h_i/2, \quad \sqrt{\varepsilon_i} \leq 2^{-i} r, \quad h_i \cdot< \Delta_i^{-1}, \quad r \mu_i \cdot< \delta_i \Delta_i^{-1}, \quad s/2 \leq s_i.$$

The last condition of (83) is satisfied, for all $i \in \mathbb{N}$, simply by the choice of Q_0 satisfying (44), as this implies (79). As for the other four conditions of (83), using the fact that the sequences ε_i, μ_i, h_i, Δ_i^{-1} and δ_i decrease at a geometric rate with respective ratio $1/16$, $1/4$, $1/2$, $1/2$ and $1/2$, it is clear that they are satisfied for any $i \in \mathbb{N}$ if and only if they are satisfied for $i = 0$. The first three conditions of (83) for $i = 0$ are nothing but (80). Moreover, using our choice of $\delta_0 = r/4$, the fourth condition of (83) for $i = 0$ reads $\mu \cdot< \Delta_0^{-1}$ and this also follows from (80).

Hence Proposition 23 can be applied, and all the conclusions of the statement follow at once from the conclusions of Proposition 23. $\square$

We can finally conclude the proof of Theorem N, by showing that one can pass to the limit $i \to +\infty$ in Proposition 24.

PROOF OF THEOREM N. Recall that we are given $\varepsilon > 0$, $\mu > 0$, $r > 0$, $s > 0$, $h > 0$ and that we define the sequences $\varepsilon_i, \mu_i, \delta_i, h_i$ in (74), and then we chose $Q_0 \geq n + 2$ satisfying (44) to define the sequences Δ_i, Q_i in (75) and σ_i in (77) and finally, $\hat{s}_i, s_i$ and r_i were defined in (78). Moreover, we have

(84) $$\begin{cases} \lim_{i \to +\infty} \varepsilon_i = \lim_{i \to +\infty} \mu_i = \lim_{i \to +\infty} h_i = 0, \\ \lim_{i \to +\infty} r_i = r - \sum_{i \in \mathbb{N}} \delta_i = r/2, \quad \lim_{i \to +\infty} s_i = s \prod_{i \in \mathbb{N}} (1 - \sigma_i)^{2n+1} \geq s/2 \end{cases}$$

and for later use, let us observe that the following series are convergent and can be made as small as one wishes thanks to condition (43) of Theorem N:

(85) $$\sum_{i=0}^{+\infty} s^{-1} C(\sigma_i) \mu_i \leq \sum_{i=0}^{+\infty} Q_i \mu_i = \sum_{i=0}^{+\infty} (\Psi(Q_i))^{-1} \Delta_i \mu_i \leq 2(\Psi(Q_0))^{-1} \Delta_0 \mu = 2 Q_0 \mu$$

(86) $$\sum_{i=0}^{+\infty} \mu_i \leq 2\mu$$

(87) $$\sum_{i=0}^{+\infty} s^{-1} C(\sigma_i) \Psi(Q_i) \mu_i \leq \sum_{i=0}^{+\infty} \Delta_i \mu_i \leq 2\Delta_0 \mu = 2 Q_0 \Psi(Q_0) \mu$$

(88) $$\sum_{i=0}^{+\infty} \Psi(Q_i) \mu_i \leq \sum_{i=0}^{+\infty} Q_i^{-1} \Delta_i \mu_i \leq 2 Q_0^{-1} \Delta_0 \mu = 2 \Psi(Q_0) \mu$$

(89) $$\sum_{i=0}^{+\infty} s^{-1} C(\sigma_i) \Psi(Q_i) \varepsilon_i \leq \sum_{i=0}^{+\infty} \Delta_i \varepsilon_i \leq 2\Delta_0 \varepsilon = 2 Q_0 \Psi(Q_0) \varepsilon.$$

Now the condition (43) of Theorem N implies that the condition (80) of Proposition 24 is satisfied; what we need to prove is that the sequences given by this Proposition 24 do converge in the Banach space of $(M, s/2)$-ultra-differentiable functions. Recall that $\mathcal{F}^0 = (E^0, F^0, G^0, \varphi^0)$ is the identity, while for $i \geq 0$,

$$(E^{i+1}, F^{i+1}, G^{i+1}, \varphi^{i+1}) = \mathcal{F}^{i+1} = \mathcal{F}^i \circ \mathcal{F}_{i+1} = (E^i, F^i, G^i, \varphi^i) \circ (E_{i+1}, F_{i+1}, G_{i+1}, \varphi_{i+1})$$

from which one easily obtains the following inductive expressions:

(90)
$$\begin{cases} E^{i+1}(\theta, \omega) = E_{i+1}(\theta, \omega) + E^i(\theta + E_{i+1}(\theta, \omega), \varphi_{i+1}(\omega)) \\ F^{i+1}(\theta, \omega) = F_{i+1}(\theta, \omega) + F^i(\theta + E_{i+1}(\theta, \omega), \varphi_{i+1}(\omega)) \cdot (\mathrm{Id} + F_{i+1}(\theta, \omega)) \\ G^{i+1}(\theta, \omega) = (F^i(\theta + E_{i+1}(\theta, \omega), \varphi_{i+1}(\omega)) + \mathrm{Id}) \cdot G_{i+1}(\theta, \omega) + G^i(\theta + E_{i+1}(\theta, \omega), \varphi_{i+1}(\omega)) \\ \varphi^{i+1} = \varphi^i \circ \varphi_{i+1}. \end{cases}$$

Let us first prove that the sequence φ^i converges. We claim that for all $i \in \mathbb{N}$, we have

$$|\nabla \varphi^i|_{s_i} \lessdot \prod_{l=0}^{i} (1 + s^{-1} C(\sigma_l) \mu_l) \lessdot 1$$

where the fact that the last product is bounded uniformly in $i \in \mathbb{N}$ follows from (85). For $i = 0$, $\varphi^0 = \mathrm{Id}$ and there is nothing to prove; for $i \in \mathbb{N}$ since $\varphi^{i+1} = \varphi^i \circ \varphi_{i+1}$ we have

$$\nabla \varphi^{i+1} = (\nabla \varphi^i \circ \varphi_{i+1}) \cdot \nabla \varphi_{i+1}$$

so that using the estimate for φ_{i+1} and $\nabla \varphi_{i+1}$ given in (81), Proposition 24, the claim follows by induction. Using this claim, and writing

$$\varphi^{i+1} - \varphi^i = \varphi^i \circ \varphi_{i+1} - \varphi^i = \left(\int_0^1 \nabla \varphi^i \circ (t\varphi_{i+1} + (1-t)\mathrm{Id}) dt \right) \cdot (\varphi_{i+1} - \mathrm{Id})$$

one finds

$$|\varphi^{i+1} - \varphi^i|_{s_{i+1}} \lessdot |\varphi_{i+1} - \mathrm{Id}|_{s_{i+1}},$$

and therefore

$$|\varphi^{i+1} - \varphi^i|_{s_{i+1}} \lessdot \mu_i.$$

Using the convergence of (84) and (86), one finds that the sequence φ^i converges to a trivial map

$$\varphi^* : \{\omega_0\} \to D_h, \quad \varphi^*(\omega_0) := \omega_0^*$$

such that

$$|\omega_0^* - \omega_0| \lessdot \mu.$$

Now let us define

$$V_{i+1}(\theta, \omega) := (\theta + E_{i+1}(\theta, \omega), \varphi_{i+1}(\omega)), \quad V_{i+1} = (\mathrm{Id} + E_{i+1}, \varphi_{i+1})$$

and observe that since $\Psi(Q_i) \geq 1$, then the estimates for E_{i+1}, ∇E_{i+1}, φ_{i+1} and $\nabla \varphi_{i+1}$ given in Proposition 24 imply that

$$|V_{i+1} - \mathrm{Id}|_{s_{i+1}} \lessdot \Psi(Q_i) \mu_i, \quad |\nabla V_{i+1} - \mathrm{Id}|_{s_{i+1}} \lessdot s^{-1} C(\sigma_i) \Psi(Q_i) \mu_i.$$

Using these estimates, and the fact that E^{i+1} can be written as

$$E^{i+1} = E_{i+1} + E^i \circ V_{i+1}$$

we can proceed as before, using the convergence of (87) to show first that

$$|\nabla E^i|_{s_i} \lessdot \sum_{l=0}^{i} s^{-1} C(\sigma_l) \Psi(Q_l) \mu_l \lessdot 1$$

and then

$$|E^{i+1} - E^i|_{s_{i+1}} \lessdot |E_{i+1}|_{s_{i+1}} \lessdot \Psi(Q_i) \mu_i.$$

Using the convergence of (84) and (88), this shows that E^i converges to a map

$$E^* : \mathbb{T}^n \times \{\omega_0\} \to \mathbb{T}^n \times D_h$$

such that
$$|E^*|_{s/2} \lessdot \Psi(Q_0)\mu.$$

For the F^i, we do have the expression
$$F^{i+1} = F_{i+1} + (F^i \circ V_{i+1}) \cdot (\mathrm{Id} + F_{i+1})$$

or alternatively
$$F^{i+1} = (\mathrm{Id} + F^i \circ V_{i+1}) \cdot F_{i+1} + F^i \circ V_{i+1}$$

and thus
$$F^{i+1} - F^i = (\mathrm{Id} + F^i \circ V_{i+1}) \cdot F_{i+1} + F^i \circ V_{i+1} - F^i.$$

As before, using the estimates on F_{i+1} and ∇F_{i+1} given in Proposition 24, one shows that

$$|\nabla F^i|_{s_i} \lessdot \sum_{l=0}^{i} s^{-2} C(\sigma_l)^2 \Psi(Q_l)\mu_l$$

but however, here, the sum above is not convergent. Yet we do have

$$sC(\sigma_i)^{-1}|\nabla F^i|_{s_i} \lessdot sC(\sigma_i)^{-1} \sum_{l=0}^{i} s^{-2} C(\sigma_l)^2 \Psi(Q_l)\mu_l \lessdot \sum_{l=0}^{i} s^{-1} C(\sigma_l) \Psi(Q_l)\mu_l \lessdot 1$$

from (88) and using the fact that the estimate for V_{i+1} can be written as
$$|V_{i+1} - \mathrm{Id}|_{s_{i+1}} \lessdot sC(\sigma_i)^{-1} s^{-1} C(\sigma_i) \Psi(Q_i)\mu_i$$

one obtains
$$|F^i \circ V_{i+1} - F^i|_{s_{i+1}} \lessdot s^{-1} C(\sigma_i) \Psi(Q_i)\mu_i.$$

By induction, one shows that

$$|F^i|_{s_i} \lessdot \sum_{l=0}^{i} s^{-1} C(\sigma_l) \Psi(Q_l)\mu_l \lessdot 1$$

from which one obtains
$$|\mathrm{Id} + F^i \circ V_{i+1}|_{s_i} \lessdot 1$$

and as a consequence,
$$|F^{i+1} - F^i|_{s_{i+1}} \lessdot s^{-1} C(\sigma_i) \Psi(Q_i)\mu_i.$$

Using the convergence of (84) and (87), this shows that F^i converges to a map
$$F^* : \mathbb{T}^n \times \{\omega_0\} \to \mathbb{T}^n \times D_h$$

such that
$$|F^*|_{s/2} \lessdot Q_0 \Psi(Q_0)\mu.$$

For G^i, we have the expression
$$G^{i+1} = (F^i \circ V_{i+1} + \mathrm{Id}) \cdot G_{i+1} + G^i \circ V_{i+1}$$

and thus
$$G^{i+1} - G^i = (F^i \circ V_{i+1} + \mathrm{Id}) \cdot G_{i+1} + G^i \circ V_{i+1} - G^i.$$

Proceeding exactly as we did for E^i and F^i, using the convergence of (84), (87) and (89), one finds that G^i converges to a map
$$G^* : \mathbb{T}^n \times \{\omega_0\} \to \mathbb{T}^n \times D_h$$

such that
$$|G^*|_{s/2} \lessdot Q_0 \Psi(Q_0)\varepsilon.$$

In summary, the map $\mathcal{F}^i$ converges to a map
$$\mathcal{F}^* : \mathbb{T}^n \times D_{r/2} \times \{\omega_0\} \to \mathbb{T}^n \times D_{r,h}$$

which belongs to $\mathcal{G}$, of the form

$$\begin{cases} \mathcal{F}^*(\theta, I, \omega_0) = (\Phi^*_{\omega_0}(\theta, I), \omega_0^*), \\ \Phi^*_{\omega_0}(\theta, I) = (\theta + E^*(\theta), I + F^*(\theta) \cdot I + G^*(\theta)) \end{cases}$$

with the estimates

(91) $\qquad |E^*|_{s/2} \lessdot \Psi(Q_0)\mu, \quad |F^*|_{s/2} \lessdot Q_0\Psi(Q_0)\mu, \quad |G^*|_{s/2} \lessdot Q_0\Psi(Q_0)\varepsilon, \quad |\omega_0^* - \omega_0| \lessdot \mu.$

Then from the estimates

$$|A^i|_{s_i} \le \varepsilon_i, \quad |B^i|_{s_i} \le \mu_i,$$

given in (82), Proposition 24, and the convergence (84), it follows that both A^i and B^i converge to zero. Next from the estimates

$$\begin{cases} |e^{i+1} - e^i \circ \varphi_{i+1}|_{s_{i+1}} \le |A^i|_{s_i}, \\ |\nabla_I^2 R^{i+1} - \nabla_I^2 R^i \circ \mathcal{F}_{i+1}|_{s_{i+1}} \lessdot \eta |F_{i+1}|_{s_{i+1}} \end{cases}$$

still given in (82), Proposition 24, one can prove in the same way as we did before, that e^i converges to a trivial map

$$e^* : \{\omega_0\} \to D_h, \quad e^*(\omega_0) := e_0^*$$

such that

(92) $$|e_0^* - e_{\omega_0^*}| \lessdot \varepsilon$$

whereas M^i converges to a map

$$M^* : \mathbb{T}^n \times D_{r/2} \times \{\omega_0\} \to \mathbb{T}^n \times D_{r,h}$$

such that, setting $R^*(\theta, I) = M^*(\theta, I)I \cdot I$,

(93) $$|\nabla_I^2 R^* - \nabla_I^2 R_{\omega_0^*}|_{s/2} \lessdot \eta Q_0 \Psi(Q_0)\mu.$$

Therefore we have

$$H \circ \mathcal{F}^*(\theta, I, \omega_0) = H_{\omega_0^*} \circ \Phi_{\omega_0}^*(\theta, I) = e_0^* + \omega_0 \cdot I + R^*(\theta, I),$$

which, together with the previous estimates (91), (92) and (93), is what we wanted to prove. $\quad \square$

5. Proof of Theorem A

In this section, we show how Theorem N implies Theorem A, following [**Pös01**].

PROOF OF THEOREM A. For simplicity, let us change the notation in ($*$) to consider a Hamiltonian $H : \mathbb{T}^n \times D \to \mathbb{R}$ of the form

$$\begin{cases} H(q,p) = h(p) + \epsilon f(q,p), \\ \nabla h(0) := \omega_0 \in \mathbb{R}^n \end{cases}$$

where H is (M, s_0)-ultra-differentiable for some $s_0 > 0$: concretely, we just replaced the variables (θ, I) by (q, p), ε by ϵ and s by s_0. Recall that we may assume that ω_0 is of the form

$$\omega_0 = (1, \bar\omega_0) \in \mathbb{R}^n, \quad \bar\omega_0 \in [-1, 1]^{n-1}.$$

For $p_0 \in B$, we expand h in a small neighborhood of p_0: writing $p = p_0 + I$ for I close to zero, we get

$$h(p) = h(p_0) + \nabla_p h(p_0) \cdot I + \int_0^1 (1-t)\nabla_p^2 h(p_0 + tI) \cdot I^2 \, dt.$$

Similarly, we expand ϵf with respect to p, in a small neighborhood of p_0:

$$\epsilon f(q, p) = \epsilon f(q, p_0) + \epsilon \nabla_p f(q, p_0) \cdot I + \epsilon \int_0^1 (1-t)\nabla_p^2 f(q, p_0 + tI) \cdot I^2 \, dt.$$

Since $\nabla_p h : B \to \Omega$ is a diffeomorphism, instead of p_0 we can use $\omega = \nabla_p h(p_0)$ as a new variable, and letting $\nabla_\omega g := (\nabla h)^{-1}$, we write

$$h(p) = e(\omega) + \omega \cdot I + R_h(I, \omega)$$

with

$$e(\omega) := h(\nabla_\omega g(\omega)), \quad R_h(I, \omega) := \int_0^1 (1-t)\nabla_p^2 h(\nabla_\omega g(\omega) + tI) \cdot I^2 \, dt$$

and also, letting $\theta = q$,

$$\epsilon f(q, p) = \epsilon \tilde{A}(\theta, \omega) + \epsilon \tilde{B}(\theta, \omega) \cdot I + \epsilon R_f(\theta, I, \omega)$$

with

$$\tilde{A}(\theta, \omega) := f(\theta, \nabla_\omega g(\omega)), \quad \tilde{B}(\theta, \omega) := \nabla_p f(\theta, \nabla_\omega g(\omega))$$

and

$$R_f(\theta, I, \omega) := \epsilon \int_0^1 (1 - t) \nabla_p^2 f(\theta, \nabla_\omega g(\omega) + tI) \cdot I^2 \, dt.$$

Finally, we can set

$$A := \epsilon \tilde{A}, \quad B := \epsilon \tilde{B}, \quad R := R_h + \epsilon R_f = M(\theta, I, \omega) \cdot I^2,$$

so that $h + \epsilon f$ can be written as

$$H(\theta, I, \omega) = e(\omega) + \omega \cdot I + A(\theta, \omega) + B(\theta, \omega) \cdot I + R(\theta, I, \omega),$$

and we have

$$\nabla_I^2 R(\theta, I, \omega) = \nabla_I^2 h(\nabla_\omega g(\omega) + I) + \epsilon \nabla_I^2 f(\theta, \nabla_\omega g(\omega) + I).$$

By assumption, h and f are (M, s_0)-ultra-differentiable on $\mathbb{T}^n \times B$, and since the space of ultra-differentiable functions is closed under taking derivatives, products, composition and inversion (by the assumptions (H1) and (H2), up to restricting the parameter s_0, see § 1 for the relevant estimates), we claim that we can find[1] $s > 0$, $r > 0$, $h > 0$ and $\tilde{c} > 0$ which are independent of ϵ such that H is (M, s)-ultra-differentiable on the domain $\mathbb{T}^n \times D_{r,h}$ with the estimates

$$|A|_s \leq \tilde{c}\epsilon, \quad |B|_s \leq \tilde{c}\epsilon, \quad |\nabla_I^2 R|_s \leq \tilde{c}.$$

We may set

$$\varepsilon := \tilde{c}\epsilon, \quad \mu := \sqrt{\varepsilon}, \quad \eta := \tilde{c}$$

and assuming ϵ small enough, we have $\tilde{c}\epsilon \leq \mu = \sqrt{\varepsilon}$. Thus we have

$$|A|_s \leq \varepsilon, \quad |B|_s \leq \mu, \quad |\nabla_I^2 R|_s \leq \eta.$$

Having fixed $s > 0$ and $r > 0$, we may choose Q_0 sufficiently large so that (44) holds true, and then by further restricting first h, and then ϵ, we may assume that the condition (43) is satisfied. Theorem N applies: there exist an $(M, s/2)$-ultra-differentiable embedding $\Upsilon_{\omega_0} : \mathbb{T}^n \to \mathbb{T}^n \times D_r$, defined by

$$\Upsilon_{\omega_0}(\theta) := \Phi_{\omega_0}(\theta, 0) = (\theta + E^*(\theta), G^*(\theta))$$

where Φ_{ω_0} is given by Theorem N, and a vector $\omega_0^* \in \mathbb{R}^n$ such that $\Upsilon_{\omega_0}(\mathbb{T}^n)$ is invariant by the Hamiltonian flow of $H_{\omega_0^*}$ and quasi-periodic with frequency ω_0. Moreover, ω_0^* and Υ_{ω_0} satisfy the estimates

$$|\omega_0^* - \omega_0| \leq c\mu, \quad |E^*|_{s/2} \leq c\Psi(Q_0)\mu, \quad |G^*|_{s/2} \leq cQ_0\Psi(Q_0)\varepsilon$$

for some large constant $c > 1$. Since h is non-degenerate, there exists p_0^* such that $\nabla h(p_0^*) = \omega_0^*$ and, up to taking $c > 1$ larger and recalling that $\mu = \sqrt{\varepsilon}$, the above estimates imply

(94) $$|p_0^*| \leq c\sqrt{\varepsilon}, \quad |E^*|_{s/2} \leq c\Psi(Q_0)\sqrt{\varepsilon}, \quad |G^*|_{s/2} \leq cQ_0\Psi(Q_0)\varepsilon.$$

Now observe that an orbit $(\theta(t), I(t))$ for the Hamiltonian $H_{\omega_0^*}$ corresponds to an orbit $(q(t), p(t)) = (\theta(t), I(t) + p_0^*)$ for our original Hamiltonian. Hence, if we define $T : \mathbb{T}^n \times \mathbb{R}^n \to \mathbb{T}^n \times \mathbb{R}^n$ by $T(\theta, I) = (\theta, I + p_0^*)$ and

$$\Theta_{\omega_0} = T \circ \Upsilon_{\omega_0} : \mathbb{T}^n \to \mathbb{T}^n \times \mathbb{R}^n, \quad \Theta_{\omega_0}(\theta) = (\theta + E^*(\theta), G^*(\theta) + p_0^*)$$

then Θ_{ω_0} is an $(M, s/2)$-ultra-differentiable torus embedding such that $\Theta_{\omega_0}(\mathbb{T}^n)$ is invariant by the Hamiltonian flow of H and quasi-periodic with frequency ω_0. The estimates on the distance between Θ_{ω_0} and the trivial embedding Θ_0 follows directly from (94), which finishes the proof. $\square$

---

[1]Beware of the local clash of notations here, since the Hamiltonian h should not be mixed up with the real parameter h.

6. Proof of Theorem F

Now we show how Theorem F follows from Theorem N.

PROOF OF THEOREM F. Consider the vector field $X = \omega_0 + B \in \mathcal{U}_s(\mathbb{T}^n, \mathbb{R}^n)$ as in the statement. It can be trivially included into a parameter-depending vector field: given $h > 0$, let $\hat{X} \in \mathcal{U}_s(\mathbb{T}^n \times D_h, \mathbb{R}^n)$ be such that

$$\hat{X}(\theta, \omega) = \hat{X}_\omega(\theta) = \omega + B(\theta), \quad \omega \in D_h, \quad \hat{X}_{\omega_0} = X.$$

Now given any $r > 0$, consider the Hamiltonian H defined on $\mathbb{T}^n \times D_{r,h}$ by

$$(95) \qquad H(\theta, I, \omega) = H_\omega(\theta, I) := \omega \cdot I + B(\theta) \cdot I.$$

Clearly, for any parameter ω, the torus $\mathbb{T}^n \times \{0\}$ is invariant by the Hamiltonian vector field X_{H_ω}, and, upon identifying $\mathbb{T}^n \times \{0\}$ with $\mathbb{T}^n$, the restriction of X_{H_ω} to this torus coincides with $\hat{X}_\omega$.

Now the Hamiltonian H defined in (95) is of the form $(**)$ with $\varepsilon = \eta = 0$ (and $e = 0$) and therefore for μ sufficiently small, Theorem N applies: there exist a vector $\omega_0^* \in \mathbb{R}^n$ and an $(M, s/2)$-ultra-differentiable embedding

$$\Phi_{\omega_0}^* : \mathbb{T}^n \times D_{r/2} \to \mathbb{T}^n \times D_r$$

here of the form

$$\Phi_{\omega_0}^*(\theta, I) = (\theta + E^*(\theta), I + F^*(\theta) \cdot I)$$

with the estimates

$$|\omega_0^* - \omega_0| \le c_3 \mu, \quad |E^*|_{s/2} \le c_3 \Psi(Q_0) \mu, \quad |F^*|_{s/2} \le c_3 \Delta(Q_0) \mu$$

and such that

$$(96) \qquad H_{\omega_0^*} \circ \Phi_{\omega_0}^*(\theta, I) = \omega_0 \cdot I.$$

The embedding $\Phi_{\omega_0}^*$ clearly leaves invariant the torus $\mathbb{T}^n \times \{0\}$ and induces a diffeomorphism of this torus that can be identified to $\Xi := \mathrm{Id} + E^*$. Writing the equality (96) in terms of Hamiltonian vector fields, we have, upon restriction to the invariant torus and recalling that the restriction of X_{H_ω} coincides with $\hat{X}_\omega$,

$$\Xi^*(\hat{X}_{\omega_0^*}) = \omega_0.$$

But $\hat{X}_{\omega_0^*} = \hat{X}_{\omega_0} + \omega_0^* - \omega_0 = X + \omega_0^* - \omega_0$ and therefore

$$\Xi^*(X + \omega_0^* - \omega_0) = \omega_0$$

which, together with the estimates on ω_0^* and $\Xi - \mathrm{Id} = E^*$, was the statement we wanted to prove. $\qquad \square$

7. Proof of Theorem B

The goal of this short section is to show how Theorem B follows directly from the work of Bessi in [**Bes00**].

In Bessi, one starts with a non-resonant vector $\omega \in \mathbb{R}^n$ which is assumed to be "exponentially Liouville" in the following sense: there exists $s_0 > 0$ and a sequence $k_j \in \mathbb{Z}^n$ with $|k_j| \to +\infty$ as $j \to +\infty$ for which

$$(C_{1,s_0}) \qquad 0 < |k_j \cdot \omega| \le e^{-s_0|k_j|}.$$

Given this sequence of $k_j \in \mathbb{Z}^n$, one can find another sequence $\tilde{k}_j \in \mathbb{Z}^n$ such that for all $j \in \mathbb{N}$, $|\tilde{k}_j| \le |k_j|$, $\tilde{k}_j \cdot k_j = 0$ and $|\tilde{k}_j \cdot \omega| \ge c|\tilde{k}_j|$ for some constant $c > 0$ independent of j.

Then one defines the following sequence of Hamiltonians on $\mathbb{R}^n/(2\pi\mathbb{Z}^n) \times \mathbb{R}^n$ (which are similar to the Hamiltonian considered by Arnold in [**Arn64**]):

$$(H_{1,j,s}) \qquad \begin{cases} H_{\varepsilon,\mu}^{1,j}(\theta, I) := \frac{1}{2} I \cdot I + F_{\varepsilon,\mu}^{1,j}(\theta), \ (\theta, I) \in \mathbb{R}^n/(2\pi\mathbb{Z}^n) \times \mathbb{R}^n \\ F_{\varepsilon,\mu}^{1,j}(\theta) := \varepsilon\nu_{1,j,s}(1 - \cos(k_j \cdot \theta))(1 + \mu\tilde{\nu}_{1,j,s}\cos(\tilde{k}_j \cdot \theta)) \\ 0 < \varepsilon \le 1, \ 0 < \mu \le 1, \ \nu_{1,j,s} := e^{-s|k_j|}, \ \tilde{\nu}_{1,j,s} := e^{-s|\tilde{k}_j|}. \end{cases}$$

Observe that the only role of the sequences $\nu_{1,j,s}$ and $\tilde{\nu}_{1,j,s}$ is to ensure that the sequence of perturbations $F_{\varepsilon,\mu}^{1,j}$ satisfy, for all $j \in \mathbb{N}$ and all $0 \le \mu \le 1$:

$$|F_{\varepsilon,\mu}^{1,j}|_s := \sup_{\theta \in \mathbb{C}^n/(2\pi\mathbb{Z}^n),\ |\mathrm{Im}(\theta)| \le s} |F_{\varepsilon,\mu}^{1,j}(\theta)| \le 4\varepsilon.$$

In [**Bes00**], Bessi proved the following theorem.

THEOREM 25 (Bessi). *Assume that $\omega \in \mathbb{R}^n$ satisfy (C_{1,s_0}). Then, if $s_0 > s$, for any $0 \le \varepsilon \le 1$, there exists $\mu_\varepsilon > 0$ and $j_\varepsilon \in \mathbb{N}$ such that for any $0 < \mu \le \mu_\varepsilon$ and any $j \ge j_\varepsilon$, the Hamiltonian system defined in $(H_{1,j,s})$ does not have any invariant torus $\mathcal{T}$ satisfying*

  (i) *$\mathcal{T}$ projects diffeomorphically to $\mathbb{T}^n$;*
  (i) *There is a C^1 diffeomorphism between $\mathbb{T}^n$ and $\mathcal{T}$ which conjugates the flow on $\mathcal{T}$ to the linear flow on $\mathbb{T}^n$ of frequency ω.*

It is clear that it is the regularity of the perturbation, here the analyticity which causes the exponential decay of the Fourier coefficients, that forces the condition (C_{1,s_0}). If the perturbation is assumed to be only of class C^r for some $r \in \mathbb{N}$, then (C_{1,s_0}) can be weakened to cover frequencies ω which are Diophantine with an exponent τ which is related to r (this can be obtained from Bessi's work; one can find a better quantitative result in [**CW13**], which also uses ideas of [**Bes00**]).

Here we would like to consider the case where the perturbation is M-ultra-differentiable, for a given sequence $M = (M_l)_{l \in \mathbb{N}}$; we will consider a slight modification of the family of Hamiltonians $(H_{1,j,s})$ to a family of Hamiltonians $(H_{M,j,s})$ depending on M, which are still analytic but for which the perturbation are bounded and small only in an M-ultra-differentiable norm.

First we need to compute the M-norm of the function $P_k(\theta) := \cos(k \cdot \theta)$ for an arbitrary $k \in \mathbb{Z}^n$. Using the fact that $(l+1)^2 \le 4^l$ for any $l \in \mathbb{N}$ and recalling the definition of the function Ω in (7), we have

$$|P_k|_s = c \sup_{l \in \mathbb{N}} \frac{(l+1)^2 s^l |k|^l}{M_l} \le c \sup_{l \in \mathbb{N}} \frac{(4|k|s)^l}{M_l} = c \exp(\Omega(4|k|s)).$$

Now we introduce the following condition on the non-resonant vector $\omega \in \mathbb{R}^n$: there exists $s_0 > 0$ and a sequence $k_j \in \mathbb{Z}^n$ with $|k_j| \to +\infty$ as $j \to +\infty$ for which

$$(C_{M,s_0}) \qquad\qquad\qquad 0 < |k_j \cdot \omega| \le \exp(-\Omega(4|k|s_0)).$$

For the sequence $M = M_1$, this condition reduces to (C_{1,s_0}), up to an unimportant factor 4. Then we consider the following modified sequence of Hamiltonians:

$$(H_{M,j,s}) \qquad \begin{cases} H_{\varepsilon,\mu}^{M,j}(\theta,I) := \frac{1}{2}I \cdot I + F_{\varepsilon,\mu}^{M,j}(\theta),\ (\theta,I) \in \mathbb{R}^n/(2\pi\mathbb{Z}^n) \times \mathbb{R}^n \\ F_{\varepsilon,\mu}^{M,j}(\theta) := \varepsilon\nu_{M,j,s}(1 - \cos(k_j \cdot \theta))(1 + \mu\tilde{\nu}_{M,j,s}\cos(\tilde{k}_j \cdot \theta)) \\ 0 < \varepsilon,\mu \le 1,\ \nu_{M,j,s} := \exp(-\Omega(4|k_j|s_0)),\ \tilde{\nu}_{M,j,s} := \exp(-\Omega(4|\tilde{k}_j|s_0)). \end{cases}$$

With these choices of $\nu_{M,j,s}$ and $\tilde{\nu}_{M,j,s}$ we have that, for all $j \in \mathbb{N}$ and all $0 \le \mu \le 1$:

$$|F_{\varepsilon,\mu}^{M,j}|_s \le 4c\varepsilon.$$

The argument of Bessi goes exactly the same of way for this family of Hamiltonians $(H_{M,j,s})$ under the condition (C_{M,s_0}), up to the following minor point one has to check: recall from (8) that Ω satisfies

$$\lim_{t \to +\infty} \frac{\Omega(t)}{\ln(t)} = +\infty$$

and thus, using the condition that $s_0 > s$, one obtains

$$\limsup_{j \to +\infty} (\exp(\Omega(4|k_j|s_0) - \Omega(4|k_j|s))) = +\infty$$

which is necessary for Bessi's argument to go through. So we have the following statement.

THEOREM 26. *Assume that $\omega \in \mathbb{R}^n$ satisfy (C_{M,s_0}). Then, if $s_0 > s$, for any $0 \le \varepsilon \le 1$, there exists $\mu_\varepsilon > 0$ and $j_\varepsilon \in \mathbb{N}$ such that for any $0 < \mu \le \mu_\varepsilon$ and any $j \ge j_\varepsilon$, the Hamiltonian system defined in $(H_{M,j,s})$ does not have any invariant torus $\mathcal{T}$ satisfying*

  (i) *$\mathcal{T}$ projects diffeomorphically to $\mathbb{T}^n$;*

(i) *There is a C^1 diffeomorphism between $\mathbb{T}^n$ and $\mathcal{T}$ which conjugates the flow on $\mathcal{T}$ to the linear flow on $\mathbb{T}^n$ of frequency ω.*

Now Theorem 26 implies Theorem B, as if ω satisfies (B_M), then it satisfies (C_{M,s_0}) for some $s_0 > 0$ and it is sufficient to consider a Hamiltonian system as in ($H_{M,j,s}$) with $s < s_0$.

Application to Hamiltonian normal forms and Nekhoroshev theory

In this section, we give applications to Hamiltonian normal forms and Nekhoroshev theory: we prove Theorem H and Theorem I, Theorem J, Theorem K and Theorem L, and finally Theorem M.

In the special case where the Hamiltonians are Gevrey regular, proofs of all those results are contained in [**MS02**], [**Bou13b**] and [**Bou11**]. Our purpose here is to give an extension to ultra-differentiable Hamiltonians; our crucial technical result will be a flexible normal form around a periodic frequency (Proposition 27 in § 1), as all the results will be deduced from it.

As before, we shall write

$$u \lessdot v \quad (\text{respectively } u \cdot\!\!\lessdot v, \quad u =\!\cdot v, \quad u \cdot\!\!= v)$$

if, for some unimportant "implicit" constant $c \geq 1$, we have

$$u \leq cv \quad (\text{respectively } cu \leq v, \quad u = cv, \quad cu = v).$$

Clearly, what we will consider as being unimportant depends on the precise setting, and it will be made clear at the beginning of each of the next sections.

1. Periodic averaging

Recall that a vector $v \in \mathbb{R}^n \setminus \{0\}$ is called periodic if there exists $t > 0$ such that $tv \in \mathbb{Z}^n$; its period T is then defined as the smallest such $t > 0$. Let v be a T-periodic vector, and let us denote by $L_v(I) = v \cdot I$ the linear integrable Hamiltonian with frequency v.

In this section, given positive real parameters $s, \delta, \mu, \rho, \nu$, we shall consider the following Hamiltonian:

$$(\sharp) \qquad \begin{cases} H(\theta, I) = L_v(I) + S(I) + R(I) + F(\theta, I), \\ H : \mathbb{T}^n \times D_\delta \to \mathbb{R} \\ |\nabla S|_s \leq \mu, \quad |\nabla R|_s \leq \rho, \quad |F|_s \leq \nu, \\ \eta := \max\{\mu, \rho\}, \end{cases}$$

where $D_\delta \subset \mathbb{R}^n$ is the ball of radius δ around the origin.

The goal of this section is to prove the following result, in which implicit constants will depend only on n and where $\{\,.\,\}$ denotes the usual Poisson bracket.

PROPOSITION 27. *Let H be as in $(\sharp)$, with M satisfying* (H1) *and* (H2) *and v T-periodic. There exists a constant $A =\!\cdot 1$ such that, for $\xi > 1$, if*

$$(97) \qquad \nu \leq C(\kappa_\xi)^{-1} s\eta, \quad C^{-1}(s(AT\eta)^{-1}) \leq \kappa_\xi, \quad \kappa_\xi := \ln \xi (24)^{-1},$$

then there exists a $(M, s/\xi)$-ultra-differentiable symplectic transformation

$$\Phi : \mathbb{T}^n \times D_{\delta/\xi} \to \mathbb{T}^n \times D_\delta$$

such that

$$H \circ \Phi = L_v + S + R + \bar{F} + \hat{F}$$

with $\{\bar{F}, L_v\} = 0$ and the estimates

$$(98) \qquad |\Phi - \mathrm{Id}|_{s/\xi} \leq 2T\nu, \quad |\bar{F}|_{s/\xi} \leq 2\nu, \quad |\hat{F}|_{s/\xi} \leq \nu \exp\left(-\kappa_\xi \left(C^{-1}(s(AT\eta)^{-1})\right)^{-1}\right).$$

Moreover, if I is an integrable Hamiltonian such that $\{I, F\} = 0$, then $\{I, \bar{F}\} = 0$.

Proposition 27 will be proved by iterating a large number of times an elementary averaging step we now describe. Let us consider the following slightly more general setting:

$$(99) \quad \begin{cases} H(\theta, I) = L_v(I) + S(I) + R(I) + G(\theta, I) + F(\theta, I), \\ H : \mathbb{T}^n \times D_\delta \to \mathbb{R} \\ |\nabla S|_s \le \mu, \quad |\nabla R|_s \le \rho, \quad |F|_s \le \nu, \quad |G| \le 2\nu, \quad \{G, L_v\} = 0, \\ \eta := \max\{\mu, \rho\} \end{cases}$$

which reduces to ($\sharp$) when $G = 0$. Given a T-periodic frequency $v \in \mathbb{R}^n \setminus \{0\}$ and a function $K : \mathbb{T}^n \times D_\delta \to \mathbb{R}$, let us define $[K]_v : \mathbb{T}^n \times D_\delta \to \mathbb{R}$ by

$$[K]_v(\theta, I) = \int_0^1 K(\theta + tTv, I)dt.$$

Clearly, $[K]_v$ is the projection of K onto the kernel of the operator $\{., L_v\}$, that is $\{K, L_v\} = 0$ if and only then $K = [K]_v$.

LEMMA 28. *Let H be as in* (99), *with M satisfying* (H1) *and* (H2) *and v T-periodic. Given $0 < \sigma < 1$, assume that*

$$(100) \quad C(\sigma)^2 T\nu \cdot < s^2.$$

Then there exist a $(M, s(1 - \sigma)^3)$-ultra-differentiable symplectic transformation

$$\Psi : \mathbb{T}^n \times D_{\delta(1-\sigma)^3} \to \mathbb{T}^n \times D_\delta$$

such that

$$H \circ \Psi = L_v + S + R + G + [F]_v + F^+$$

with the estimates

$$(101) \quad |\Psi - \mathrm{Id}|_{s(1-\sigma)^3} \le T\nu, \quad |F^+|_{s(1-\sigma)^3} \lessdot (T\nu^2 C(\sigma)^2 s^{-2} + T\eta\nu C(\sigma)s^{-1}).$$

Moreover, if I is an integrable Hamiltonian such that $\{I, F\} = \{I, G\} = 0$, then $\{I, [F]_v\} = \{I, F^+\} = 0$.

PROOF. Let us define $Y : \mathbb{T}^n \times D_\delta \to \mathbb{R}$ by

$$Y(\theta, I) = T \int_0^1 (F - [F]_v)(\theta + tTv, I)t \, dt.$$

It is easy to check, using an integration by part, that

$$(102) \quad \{Y, N_v\} = F - [F]_v = F + S + R + G - [F + S + R + G]_v$$

where the second equality follows from the fact that $[S + R + G]_v = S + R + G$ since S, R are integrable and $\{G, L_v\} = 0$ by assumption. Moreover, we have the estimates

$$(103) \quad |Y|_s \le \frac{T}{2}|F - [F]_v|_s \le T|F|_s \le T\nu.$$

Recalling from (6) that $(e\sigma)^{-1} \le C(\sigma)$, the inequalities (100) imply that (37) is satisfied, and thus Proposition 19 can be applied: for any $t \in [0, 1]$, the time-t map Y^t of the Hamiltonian flow of Y belongs to $\mathcal{U}_{s(1-\sigma)^3}(\mathbb{T}^n \times D_{\delta(1-\sigma)^3}, \mathbb{T}^n \times D_{\delta(1-\sigma)^2})$ and we have the estimate

$$(104) \quad |Y^t - \mathrm{Id}|_{s(1-\sigma)^3} \le t|Y|_s.$$

We set $\Psi = Y^1$: then (104) and (103) gives the first estimate of (101). Now a well-known computation using Taylor's formula with integral remainder yields

$$H \circ \Psi = L_v + S + R + G + [F]_v + F^+$$

with

$$F^+ := \int_0^1 \tilde{F}_t \circ Y^t dt, \quad \tilde{F}_t := \{tF + (1-t)[F]_v + S + R + G, Y\}.$$

To estimate F^+, observe that Y^t maps $\mathbb{T}^n \times D_{\delta(1-\sigma)^3}$ into $\mathbb{T}^n \times D_{\delta(1-\sigma)^2}$, so using (104) and (100), we can apply Proposition 16 to get

$$|F^+|_{s(1-\sigma)^3} \leq |\tilde{F}_t|_{s(1-\sigma)^2}.$$

Now $\tilde{F}_t$, for any $t \in [0,1]$, can be easily estimated using Corollary 9 and thus we get

$$
\begin{aligned}
|\tilde{F}_t|_{s(1-\sigma)^2} \quad &<\quad (T\nu^2 C(\sigma)^2 s^{-2} + T\nu^2 C(\sigma)^2 s^{-2} + T\nu\mu C(\sigma)s^{-1} \\
&\qquad + T\nu\rho C(\sigma)s^{-1} + T\nu^2 C(\sigma)^2 s^{-2}) \\
&<\quad (T\nu^2 C(\sigma)^2 s^{-2} + T\nu\eta C(\sigma)s^{-1})
\end{aligned}
$$

as $\eta = \max\{\mu, \rho\}$ by definition. This gives the second estimate of (101).

It remains to prove the second part of the statement, so let I be an integrable Hamiltonian such that $\{I, F\} = \{I, G\} = 0$. Obviously, as I is integrable we also have $\{I, L_v\} = 0$, which gives $\{I, [F]_v\} = 0$: indeed, letting L_v^{tT} the tT-time of the Hamiltonian flow associated to L_v, as $\{I, L_v\} = 0$ we have $I \circ L_v^{tT} = I$ hence

$$
\begin{aligned}
\{I, [F]_v\} \quad &=\quad \left\{ I, \int_0^1 F \circ L_v^{tT} dt \right\} \\
&=\quad \int_0^1 \{I, F \circ L_v^{tT}\} dt \\
&=\quad \int_0^1 \{I \circ L_v^{tT}, F \circ L_v^{tT}\} dt \\
&=\quad \int_0^1 \{I, F\} \circ L_v^{tT} dt \\
&=\quad 0.
\end{aligned}
$$

Using this, one obtains then $\{I, Y\} = 0$ by a similar argument. Now using $\{I, Y\} = 0$ and again the same argument, to show that $\{I, F^+\} = 0$ one needs to show that $\{I, \tilde{F}_t\} = 0$ for all $t \in [0,1]$: but this in turns follows from Jacobi identity, as each element in the bracket $\tilde{F}_t$ commutes with I. This concludes the proof. $\qquad\square$

To prove Proposition 27, it remains to apply inductively Lemma 28 m times, for some optimized choice of m, choosing the same $\sigma \sim m^{-1}$ at each step. Actually, to obtain sharper estimates, we will use a trick due to Neishtadt which consists in applying Lemma 28 a first time with $\sigma \sim 1$ large, and then $(m-1)$ times with $\sigma \sim m^{-1}$ uniformly small.

PROOF OF PROPOSITION 27. Let $m \geq 1$ be an integer to be chosen below. We set $\sigma_1 := \ln\xi(24)^{-1}$ and for $m \geq 2$, $\sigma_m := \ln\xi(24(m-1))^{-1}$. For all $0 \leq j \leq m$, we define

$$\nu_j := e^{-j}\nu, \quad \gamma_j := \sum_{i=0}^{j-1} \nu_i,$$

and

$$s_j := s(1-\sigma_1)^3(1-\sigma_m)^{3(j-1)}, \quad \delta_j := \delta(1-\sigma_1)^3(1-\sigma_m)^{3(j-1)}.$$

We claim that if

(105) $$\nu \leq C(\sigma_1)^{-1}s\eta, \quad T\eta \cdot < C(\sigma_m)^{-1}s$$

then for all $1 \leq j \leq m$, there exists a (M, s_j)-ultra-differentiable symplectic transformation

$$\Phi_j : \mathbb{T}^n \times D_{\delta_j} \to \mathbb{T}^n \times D_\delta$$

such that

$$H \circ \Phi_j = L_v + S + R + \bar{F}_j + \hat{F}_j$$

with $\{\bar{F}_j, L_v\} = 0$ and the estimates

$$|\Phi_j - \mathrm{Id}|_{s_j} \leq T\gamma_j, \quad |\bar{F}_j|_{s_j} \leq \gamma_j, \quad |\hat{F}_j|_{s_j} < \nu_{j-1}(T\eta C(\sigma_1)s^{-1}) := \chi_j.$$

Moreover, if I is an integrable Hamiltonian such that $\{I, F\} = 0$, then we have $\{I, \bar{F}_j\} = \{I, \hat{F}_j\} = 0$. Observe that $\gamma_j < 2\nu$ and as $\sigma_1 \geq \sigma_m$, one can use (105) to ensure that $\chi_j \leq \nu_j$.

The statement follows from this claim: indeed, it suffices to choose the integer m as

$$\ln \xi(24)^{-1} \left(C^{-1}(s(AT\eta)^{-1})\right)^{-1} \leq m \leq 1 + \ln \xi(24)^{-1} \left(C^{-1}(s(AT\eta)^{-1})\right)^{-1}$$

for a suitable constant $A = \cdot 1$. Clearly (97) gives (105) and $m \geq 1$, so we may set

$$\Phi := \Phi_m, \quad \bar{F} := \bar{F}_m, \quad \hat{F} := \hat{F}_m,$$

observe that $s_m \geq s/\xi$, $\delta_m \geq \delta/\xi$, and thus

$$\Phi : \mathbb{T}^n \times D_{\delta/\xi} \to \mathbb{T}^n \times D_\delta$$

is a $(M, s/\xi)$-ultra-differentiable symplectic transformation such that

$$H \circ \Phi = L_v + S + R + \bar{F} + \hat{F}$$

with $\{\bar{F}, L_v\} = 0$ and the estimates

$$|\Phi - \mathrm{Id}|_{s/\xi} \leq 2T\nu, \quad |\bar{F}|_{s/\xi} \leq 2\nu,$$

and

$$|\hat{F}|_{s/\xi} \leq \nu \exp(-m) \leq \nu \exp\left(-\ln \xi(24)^{-1} \left(C^{-1}(s(AT\eta)^{-1})\right)^{-1}\right).$$

So it remains to prove the claim, and we will do it by induction on j. The second part of the claim is obvious by induction using the second part of Lemma 28, so we will only prove the first part of the claim. For $j = 1$, (105) implies that

$$C(\sigma_1)^2 T\nu \leq C(\sigma_1)T\eta s \cdot < s^2$$

and thus (100) is satisfied: Lemma 28 applies, with $\sigma = \sigma_1$ and $G = 0$, and we set $\Phi_1 := \Psi$, $\bar{F}_1 := [F]_v$ and $\hat{F}_1 = F^+$. Clearly, $\{\bar{F}_1, L_v\} = 0$. As $\gamma_1 = \nu_0 = \nu$, we do have

$$|\Phi_1 - \mathrm{Id}|_{s_1} \leq T\gamma_1, \quad |\bar{F}_1|_{s_1} \leq \gamma_1.$$

For $\hat{F}_1$, observe that from the first part of (105) one has

$$T\nu^2 C(\sigma_1)^2 s^{-2} \leq T\nu\eta C(\sigma_1)s^{-1}$$

and thus

$$|\hat{F}_1|_{s_1} < (T\nu^2 C(\sigma_1)^2 s^{-2} + T\nu\eta C(\sigma_1)s^{-1}) < T\nu\eta C(\sigma_1)s^{-1} = \chi_1.$$

Now assume that the statement is true for some $1 \leq j \leq m - 1$, and let us prove it remains true for $2 \leq j + 1 \leq m$. The inductive hypothesis gives a (M, s_j)-ultra-differentiable symplectic transformation

$$\Phi_j : \mathbb{T}^n \times D_{\delta_j} \to \mathbb{T}^n \times D_\delta$$

such that

$$H \circ \Phi_j = L_v + S + R + \bar{F}_j + \hat{F}_j$$

with $\{\bar{F}_j, L_v\} = 0$ and the estimates

$$|\Phi_j - \mathrm{Id}|_{s_j} \leq T\gamma_j, \quad |\bar{F}_j|_{s_j} \leq \gamma_j, \quad |\hat{F}_j|_{s_j} < \nu_{j-1}(T\eta C(\sigma_1)s^{-1}).$$

We wish to apply Lemma 28 to this Hamiltonian with $\sigma = \sigma_m$, $\bar{F}_j$ instead of G and $\hat{F}_j$ instead of F: we have

$$|\bar{F}_j| \leq \gamma_j \leq 2\nu$$

and, from the first part of (105)

$$\tag{106} T\chi_j \leq T\nu_{j-1}T\mu C(\sigma_1)s^{-1} \leq T\nu T\eta C(\sigma_1)s^{-1} \leq (T\eta)^2$$

so that the second part of (105) implies (100). Lemma 28 applies to give a transformation Ψ_j such that, if define $\Phi_{j+1} := \Phi_j \circ \Psi_j$, then

$$\Phi_{j+1} : \mathbb{T}^n \times D_{\delta_{j+1}} \to \mathbb{T}^n \times D_\delta$$

is an (M, s_{j+1})-ultra-differentiable symplectic transformation such that

$$H \circ \Phi_{j+1} = L_v + S + R + \bar{F}_j + [\hat{F}_j]_v + \hat{F}_j^+$$

and hence, setting $\bar{F}_{j+1} := \bar{F}_j + [\hat{F}_j]_v$ and $\hat{F}_{j+1} := \hat{F}_j^+$, we have

$$H \circ \Phi_{j+1} = L_v + S + R + \bar{F}_{j+1} + \hat{F}_{j+1}.$$

Clearly $\{\bar{F}_{j+1}, L_v\} = 0$ and as $|\hat{F}_j|_{s_j} \le \nu_j$, we have

$$|\bar{F}_{j+1}|_{s_{j+1}} \le |\bar{F}_j|_{s_j} + |[\hat{F}_j]_v|_{s_j} \le \gamma_j + \nu_j = \gamma_{j+1}$$

and, using Proposition 16,

$$|\Phi_{j+1} - \mathrm{Id}|_{s_{j+1}} \le |\Phi_j - \mathrm{Id}|_{s_j} + |\Psi_j - \mathrm{Id}|_{s_{j+1}} \le T\gamma_j + T\nu_j = T\gamma_{j+1}.$$

For $\hat{F}_{j+1}$, using (106) and the second part of (105) we can estimate

$$
\begin{aligned}
|\hat{F}_{j+1}| \;&<\; (T\chi_j^2 C(\sigma_m)^2 s^{-2} + T\eta\chi_j C(\sigma_m) s^{-1}) \\
&<\; \chi_j (T^2\eta^2 C(\sigma_m)^2 s^{-2} + T\eta C(\sigma_m) s^{-1}) \\
&<\; \chi_j T\eta C(\sigma_m) s^{-1} \\
&\le\; \chi_{j+1}.
\end{aligned}
$$

This proves the claim, and thus the proposition. $\qquad\square$

2. Stability in the linear case: proof of Theorem H

The goal here is to prove Theorem H and Corollary 1; implicit constants will depend only on n, ω and the function C.

Our proof will use Proposition 27 (which can be seen as Theorem H in the special case where $d = 1$) which we proved in § 1. We will also use a slightly more general version of Proposition 21 (§ 2) which we now state and which is still contained in [**BF13**].

PROPOSITION 29. *Let $\omega = (\bar{\omega}, 0) \in \mathbb{R}^d \times \mathbb{R}^{n-d}$ with $\bar{\omega}$ a non-resonant vector. For any $Q \ge n + 2$, there exist d periodic vectors $v_1, \ldots, v_d \in \mathbb{R}^d \times \{0\}$, of periods $T_1, \ldots, T_d$, such that $T_1 v_1, \ldots, T_d v_d$ form a $\mathbb{Z}$-basis of $\mathbb{Z}^d \times \{0\}$ and for $j \in \{1, \ldots, d\}$,*

$$|\omega_0 - v_j| < (T_j Q)^{-1}, \quad 1 \le T_j < \Psi(Q).$$

Given any $\omega \in \mathbb{R}^n$, let us now define

$$(107) \qquad [H]_\omega(\theta, I) := \lim_{s \to +\infty} \frac{1}{s} \int_0^s H(\theta + t\omega, I)\,dt.$$

When $\omega = (\bar{\omega}, 0) \in \mathbb{R}^d \times \mathbb{R}^{n-d}$ with $\bar{\omega}$ a non-resonant vector, clearly $[H]_\omega = [H]_d$ where

$$[H]_d(\theta, I) = \int_{\mathbb{T}^d} H(\theta, I)\,d\theta_1 \ldots d\theta_d.$$

We then have the following result, which, as before, is slightly more general than Proposition 30, and is a consequence of the fact that $T_1 v_1, \ldots, T_d v_d$ form a $\mathbb{Z}$-basis of $\mathbb{Z}^d \times \{0\}$.

PROPOSITION 30. *Let $v_1, \ldots, v_d \in \mathbb{R}^d \times \{0\}$ be the periodic vectors given by Proposition 29, and H a function defined on $\mathbb{T}^n \times D_r$. Then*

$$[H]_{v_1, \ldots, v_d} = [H]_\omega = [H]_d$$

and $\{H, L_\omega\} = 0$ if and only if $\{H, L_{v_j}\} = 0$ for any $1 \le j \le d$.

Given positive real parameters s, ρ, ν, we shall consider the following Hamiltonian:

$$(108) \qquad \begin{cases} H(\theta, I) = L_\omega(I) + R(I) + F(\theta, I), \\ H : \mathbb{T}^n \times D \to \mathbb{R}, \quad \omega = (\bar{\omega}, 0) \in \mathbb{R}^d \times \mathbb{R}^{n-d} \\ |\nabla R|_s \le \rho, \quad |F|_s \le \nu \end{cases}$$

where $\bar{\omega}$ is non-resonant. In the special case where $R = 0$, $\rho = 0$ and $\nu = \varepsilon$, this is exactly a Hamiltonian H as in $(*)$ with $h = L_\omega$, which is the setting of Theorem H.

We are now ready to prove the following statement, which will easily imply Theorem H.

PROPOSITION 31. *Let H be as in (108), where H is (M,s)-ultra-differentiable with M satis-fying (H1) and (H2) and $h = L_\omega$ with $\omega = (\bar\omega, 0) \in \mathbb{R}^d \times \{0\}$ and $\bar\omega$ non-resonant. For $Q \geq n+2$, assume that*

$$(109) \qquad \rho Q \Psi(Q) \cdot< 1, \quad \nu Q \Psi(Q) s^{-1} \cdot< 1, \quad \lambda s Q > \cdot 1,$$

for some $\lambda \cdot= 1$. Then there exists a $(M, s/2)$-ultra-differentiable symplectic transformation

$$\Phi : \mathbb{T}^n \times D_{1/2} \to \mathbb{T}^n \times D$$

such that

$$H \circ \Phi = L_\omega + R + \bar{F} + \hat{F}$$

where $\{\bar{F}, L_\omega\} = 0$ and with the estimates

$$|\Phi - \mathrm{Id}|_{s/2} \cdot< \Psi(Q)\nu, \quad |\bar{F}|_{s/2} \cdot< \nu, \quad |\hat{F}|_{s/2} \cdot< \nu \exp\left(-\kappa_{2^{1/d}}\left(C^{-1}(\lambda s Q)\right)^{-1}\right).$$

PROOF OF THEOREM H. Here $R = 0$, $\rho = 0$ and $F = f$ with $\nu = \varepsilon$, it suffices to choose

$$Q := \Delta^{-1}\left(c_2 s \varepsilon^{-1}\right), \quad c_2 = \cdot 1$$

and to verify that choosing ε sufficiently small allows indeed to apply Proposition 31. $\qquad \square$

PROOF OF PROPOSITION 31. Recall that we are considering

$$H = L_\omega + R + F.$$

Since $Q \geq n+2$, we can apply Proposition 29: there exist d periodic vectors $v_1, \ldots, v_d$, of periods $T_1, \ldots, T_d$, such that $T_1 v_1, \ldots, T_d v_d$ form a $\mathbb{Z}$-basis of $\mathbb{Z}^d \times \{0\}$ and for $j \in \{1, \ldots, d\}$,

$$|\omega - v_j| \cdot< (T_j Q)^{-1}, \quad 1 \cdot< T_j \cdot< \Psi_\omega(Q).$$

For $j \in \{1, \ldots, d\}$, let us define

$$S_j = L_\omega - L_{v_j}, \quad \mu_j = \cdot (T_j Q)^{-1}$$

with a suitable implicit constant so that

$$|\nabla S_j|_s = c|\omega - v_j| \leq \mu_j.$$

Note that $L_\omega = L_{v_j} + S_j$, and that $T_j \nu_j = \cdot Q^{-1}$ for any $1 \leq j \leq d$. Let us further choose $\xi := 2^{1/d}$ and define, for $1 \leq j \leq d$

$$s_j := s\xi^{-j}, \quad \delta_j := \xi^{-j}$$

so that in particular $s_d = s/2$ and $\delta_d = 1/2$.

We claim that for all $1 \leq j \leq d$, there exists a (M, s_j)-ultra-differentiable symplectic map

$$\Phi_j : \mathbb{T}^n \times D_{\delta_j} \longrightarrow \mathbb{T}^n \times D$$

such that

$$H \circ \Phi_j = L_\omega + R + \bar{F}_j + \hat{F}_j,$$

with $\{\bar{F}_j, L_{v_i}\} = 0$ for any $1 \leq i \leq j$, and with the estimates

$$(110) \qquad \begin{cases} |\Phi_j - \mathrm{Id}|_{s_j} \cdot< \Psi(Q)\nu, \quad |\bar{F}_j|_{s_j} \cdot< \nu, \\ |\hat{F}_j|_{s_j} \cdot< \nu \exp(-\kappa_{2^{1/d}}\left(C^{-1}(\lambda s Q)\right)^{-1}). \end{cases}$$

The proof of the proposition follows from this claim: it is sufficient to let

$$\Phi := \Phi_d, \quad \bar{F} := \bar{F}_d, \quad \hat{F} := \hat{F}_d$$

since from Corollary 22, $\{\bar{F}, L_{v_i}\} = 0$ for $1 \leq i \leq d$ is equivalent to $\{\bar{F}, L_\omega\} = 0$.

Now let us prove the claim by induction on $1 \leq j \leq d$. For $j = 1$, our Hamiltonian can be written as

$$H = L_\omega + R + F = L_{v_1} + S_1 + R + F$$

and this is nothing but Proposition 27, which can be applied with μ_1 instead of μ, because (109) implies that $\eta = \max\{\mu_1, \rho\} = \mu_1$ and (97): indeed, since $T_1 \cdot< \Psi(Q)$ and $T_1 \mu_1 = \cdot Q^{-1}$, the inequalities (109) imply

$$\rho \leq \mu_1, \quad \nu \cdot< s\eta, \quad T\eta \cdot< s$$

which gives (97) for a proper choice of implicit constants. So now assume that the statement holds true for some $1 \leq j \leq d-1$, and let us prove that it is true for $2 \leq j+1 \leq d$. By the inductive assumption, there exists a (M, s_j)-ultra-differentiable symplectic map

$$\Phi_j : \mathbb{T}^n \times D_{\delta_j} \longrightarrow \mathbb{T}^n \times D$$

such that

$$H \circ \Phi_j = L_\omega + R + \bar{F}_j + \hat{F}_j,$$

with $\{\bar{F}_j, L_{v_i}\} = 0$ for any $1 \leq i \leq j$, and with the estimates

$$(111) \qquad \begin{cases} |\Phi_j - \mathrm{Id}|_{s_j} \lessdot \Psi(Q)\nu, \quad |\bar{F}_j|_{s_j} \lessdot \nu, \\ |\hat{F}_j|_{s_j} \lessdot \nu \exp(-\kappa_{2^{1/d}} \left(C^{-1}(\lambda s Q) \right)^{-1}). \end{cases}$$

Let us consider the Hamiltonian

$$\bar{H}_j := H \circ \Phi_j - \hat{F}_j = L_\omega + R + \bar{F}_j$$

which is (M, s_j)-ultra-differentiable on the domain $\mathbb{T}^n \times D_{\delta_j}$. It can also be written as

$$\bar{H}_j = L_{v_{j+1}} + S_{j+1} + R + \bar{F}_j$$

and because of (109), we can still apply Proposition 27 to this Hamiltonian with μ_{j+1} and a constant times ν instead of respectively μ and ν (and as before, $\eta = \mu_{j+1}$). Therefore, since $s_j/\xi = s_{j+1}$ and $\delta_j/\xi = \delta_{j+1}$, we can find a (M, s_{j+1})-ultra-differentiable symplectic transformation

$$\bar{\Phi} : \mathbb{T}^n \times D_{\delta_{j+1}} \to \mathbb{T}^n \times D_{\delta_j}$$

such that

$$\bar{H}_j \circ \bar{\Phi} = L_\omega + R + \bar{\bar{F}}_j + \widehat{\bar{F}}_j$$

with $\{\bar{\bar{F}}_j, L_{v_{j+1}}\} = 0$ and the estimates

$$|\bar{\Phi} - \mathrm{Id}|_{s_{j+1}} \lessdot T_{j+1}\nu, \quad |\bar{\bar{F}}_j|_{s_{j+1}} \lessdot \nu, \quad |\widehat{\bar{F}}_j|_{s_{j+1}} \lessdot \nu \exp(-\kappa_{2^{1/d}} \left(C^{-1}(\lambda s Q) \right)^{-1}).$$

Moreover, for $1 \leq i \leq j$, since L_{v_i} are integrable Hamiltonians for which $\{L_{v_i}, \bar{F}_j\} = 0$, we have $\{L_{v_i}, \bar{\bar{F}}_j\} = 0$. Hence $\{\bar{\bar{F}}_j, L_{v_i}\} = 0$ for $1 \leq i \leq j+1$. We may set

$$\Phi_{j+1} := \Phi_j \circ \bar{\Phi}, \quad \bar{F}_{j+1} := \bar{\bar{F}}_j, \quad \hat{F}_{j+1} := \widehat{\bar{F}}_j + \hat{F}_j \circ \bar{\Phi}$$

so that

$$H \circ \Phi_{j+1} = L_\omega + R + \bar{F}_{j+1} + \hat{F}_{j+1}.$$

We already explained that $\{\bar{F}_{j+1}, L_{v_i}\} = 0$ for $1 \leq i \leq j+1$, and

$$|\bar{F}_{j+1}|_{s_{j+1}} \lessdot \nu.$$

It is obvious that

$$\Phi_{j+1} : \mathbb{T}^n \times D_{\delta_{j+1}} \to \mathbb{T}^n \times D_\delta$$

and using Proposition 16 and the fact that $T_{j+1} \lessdot \Psi(Q)$, one has

$$|\Phi_{j+1} - \mathrm{Id}|_{s_{j+1}} \leq |\Phi_{j+1} - \mathrm{Id}|_{s_j} + |\bar{\Phi} - \mathrm{Id}|_{s_{j+1}} \lessdot \Psi(Q)\nu.$$

To conclude, using once again Proposition 16 we have

$$|\hat{F}_{j+1}|_{s_{j+1}} \leq |\widehat{\bar{F}}_j|_{s_{j+1}} + |\hat{F}_j|_{s_j} \lessdot \nu \exp(-\kappa_{2^{1/d}} \left(C^{-1}(\lambda s Q) \right)^{-1}).$$

This proves that the statement holds true for $j+1$, which finishes the induction and the proof. $\square$

Let us conclude by giving the proof of Corollary 1.

Proof of Corollary 1. Since we are under the assumptions of Theorem H, let us first consider the Hamiltonian in normal form

$$H \circ \Phi = h + \bar{f} + \hat{f}$$

where

$$\Phi : \mathbb{T}^n \times D_{1/2} \to \mathbb{T}^n \times D,$$

$\{\bar{f}, L_\omega\} = 0$ and with the estimates

$$\begin{cases} |\Phi - \mathrm{Id}|_{s/2} \leq c_1 \Psi(\Delta^{-1}(c_2 s \varepsilon^{-1}))\varepsilon, \quad |\bar{f}|_{s/2} \leq c_1 \varepsilon, \\ |\hat{f}|_{s/2} \leq c_1 \varepsilon \exp(-c_3 \left(C^{-1}(c_4 s \Delta^{-1}(c_2 s \varepsilon^{-1}))\right)^{-1}). \end{cases}$$

Let $(\tilde{\theta}(t), \tilde{I}(t))$ be a solution of $H \circ \Phi$ with $\tilde{I}(0) \in D_{1/4}$. Since $\{\bar{f}, L_\omega\} = 0$ and $\omega = (\bar{\omega}, 0) \in \mathbb{R}^d \times \mathbb{R}^{n-d}$ with $\bar{\omega}$ non-resonant, we have $\partial_{\theta_j} \bar{f} = 0$ for $1 \leq j \leq d$ and thus

$$|\Pi_d(\tilde{I}(t) - \tilde{I}(0))| \leq |t| \sup_{(\tilde{\theta}, \tilde{I}) \in \mathbb{T}^n \times D_{1/2}} |\partial_\theta \hat{f}(\tilde{\theta}, \tilde{I})|$$

for all t as long as $\tilde{I}(t) \in D_{1/4}$. By Corollary 9 and the above estimate on $\hat{f}$, we have

$$\sup_{(\tilde{\theta}, \tilde{I}) \in \mathbb{T}^n \times D_{1/2}} |\partial_\theta \hat{f}(\tilde{\theta}, \tilde{I})| \lessdot s^{-1} \varepsilon \exp(-c_3 \left(C^{-1}(c_4 s \Delta^{-1}(c_2 s \varepsilon^{-1}))\right)^{-1})$$

hence if we define

$$T \cdot = r s \varepsilon^{-1} \exp(c_3 \left(C^{-1}(c_4 s \Delta^{-1}(c_2 s \varepsilon^{-1}))\right)^{-1})$$

we obtain

$$|\Pi_d(\tilde{I}(t) - \tilde{I}(0))| \leq r, \quad |t| \leq T$$

since $r \leq 1/4$ and as long as $\tilde{I}(t) \in D_{1/2}$. Now coming back to the original Hamiltonian H, any solution $(\theta(t), I(t))$ with $I(0) \in D_{1/8}$ gives rise to a solution $(\tilde{\theta}(t), \tilde{I}(t)) = \Phi^{-1}(\theta(t), I(t))$ of $H \circ \Phi$ with $\tilde{I}(0) \in 1/4$, and therefore, for $|t| \leq T$ and such that $I(t) \in D_{1/4}$, we obtain

$$\begin{aligned} |\Pi_d(I(t) - I(0))| &\leq |\Pi_d(I(t) - \tilde{I}(t))| + |\Pi_d(\tilde{I}(t) - \tilde{I}(0))| + |\Pi_d(\tilde{I}(0) - I(0))| \\ &\leq r + 2c_1 \Psi(\Delta^{-1}(c_2 s \varepsilon^{-1}))\varepsilon \\ &\leq 2r \end{aligned}$$

since we assumed $r \geq 2c_1 \Psi(\Delta^{-1}(c_2 s \varepsilon^{-1}))\varepsilon$. This concludes the proof. $\qquad \square$

3. Diffusion in the linear case: proof of Theorem I

In this section we shall give the proof of Theorem I.

Proof. For convenience, let us write

$$\omega = (\bar{\omega}, 0) = (1, \bar{\omega}_1, \dots, \bar{\omega}_{d-1}, 0) \in \mathbb{R}^d \times \mathbb{R}^{n-d}$$

with $|\bar{\omega}_i| \leq 1$ for $1 \leq i \leq d-1$. Let us denote by $(p_j/q_j)_{j \in \mathbb{N}}$ the sequence of the convergents of $\bar{\omega}_1$ for instance. We have the classical inequalities

$$(q_j + q_{j+1})^{-1} < |q_j \bar{\omega}_1 - p_j| < q_{j+1}^{-1}, \quad j \in \mathbb{N},$$

and since $q_{j+1} > q_j$, this gives

$$(112) \qquad (2q_{j+1})^{-1} < |q_j \bar{\omega}_1 - p_j| < q_{j+1}^{-1}, \quad j \in \mathbb{N}.$$

Now by definition of Ψ_ω, we obtain

$$(113) \qquad q_{j+1} < \Psi_\omega(q_j) < 2q_{j+1}.$$

The perturbation f_j will be of the form

$$f_j(\theta, I) = f_j^1(I) + f_j^2(\theta), \quad (\theta, I) \in \mathbb{T}^n \times \mathbb{R}^n.$$

First, we choose $f_j^1(I) = v_j \cdot I - \omega \cdot I$, where $v_j = (1, p_j/q_j, \bar{\omega}_2, \dots, \bar{\omega}_{d-1}, 0)$. We set

$$\varepsilon_j = |\bar{\omega}_1 - p_j/q_j|$$

and observe that ε_j tends to zero as j tends to infinity. From the inequalities (112) and (113), recalling the definitions of Δ_ω and Δ_ω^*, we have

$$(114) \qquad (2\Delta_\omega(q_j))^{-1} \leq \varepsilon_j \leq 2(\Delta_\omega(q_j))^{-1}, \quad \Delta_\omega^*((2\varepsilon_j)^{-1}) \leq q_j \leq \Delta_\omega^*(2\varepsilon_j^{-1}).$$

Now by definition, $|\partial_I f_j^1|_s = c\varepsilon_j$. Then, if we let $k_j = (p_j, -q_j, 0, \ldots, 0) \in \mathbb{Z}^n$, we define $f_j^2(\theta) = \varepsilon_j \mu_j \sin(2\pi k_j \cdot \theta)$ with μ_j to be chosen. Note that

$$(115) \qquad q_j \leq |k_j| \leq 2q_j$$

since $|q_j| \geq |p_j|$ (as $|\bar{\omega}_1| \leq 1$). It is easy to estimate

$$|\partial_\theta f_j^2|_s \leq c\varepsilon_j \mu_j 2\pi |k_j| \exp\left(\Omega(8\pi|k_j|s)\right)$$

and so we choose

$$\mu_j = (2\pi|k_j|)^{-1} \exp\left(-\Omega(8\pi|k_j|s)\right)$$

so that $|\partial_\theta f_j^2|_s \leq c\varepsilon_j$. Finally, we set $f_j = f_j^1 + f_j^2$ and we have

$$|X_{f_j}|_s \leq |\partial_I f_j^1|_s + |\partial_\theta f_j^2|_s \leq 2c\varepsilon_j.$$

Now we can write the Hamiltonian

$$\begin{aligned} H_j(\theta, I) &= L_\omega(I) + f_j(\theta, I) \\ &= \omega \cdot I + v_j \cdot I - \omega \cdot I + \varepsilon_j \mu_j \sin(2\pi k_j \cdot \theta) \\ &= v_j \cdot I + \varepsilon_j \mu_j \sin(2\pi k_j \cdot \theta) \end{aligned}$$

and as $k_j \cdot v_j = 0$, the associated system is easily integrated:

$$\begin{cases} \theta(t) &= \theta_0 + tv_j \quad [\mathbb{Z}^n] \\ I(t) &= I_0 - t2\pi k_j \varepsilon_j \mu_j \cos(2\pi k_j.\theta_0). \end{cases}$$

Choosing any solution $(\theta(t), I(t))$ with an initial condition $(\theta(0), I(0))$ for which $k_j \cdot \theta(0)$ is an integer, using the definition of μ_j, we obtain

$$|I(t) - I(0)| = |I_1(t) - I_1(0)| + |I_2(t) - I_2(0)| = |t|\varepsilon_j \exp\left(-\Omega(8\pi|k_j|s)\right).$$

Using (115) and (114) this gives

$$|t|\varepsilon_j \exp\left(-\Omega(16\pi s\Delta_\omega^*(2\varepsilon_j^{-1}))\right) \leq |I(t) - I(0)| \leq |t|\varepsilon_j \exp\left(-\Omega(8\pi s\Delta_\omega^*((2\varepsilon_j)^{-1}))\right)$$

which is what we wanted to prove, as $|I(t) - I(0)| = |\Pi_d(I(t) - I(0))|$. $\qquad\square$

4. Stability in the non-linear case: proof of Theorem J

The goal here is to prove Theorem J and Corollary 3; implicit constants will depend only on n, ω, $|h|_s$ and the function C.

As before, Theorem J will be an easy consequence of the following proposition, where we recall that given $\rho > 0$, we let

$$\sigma = \sigma_\rho : (\theta, I) \longmapsto (\theta, \rho I).$$

PROPOSITION 32. *Let H be as in $(*)$, where H is (M, s)-ultra-differentiable with M satisfying* (H1) *and* (H2) *and $\omega = (\bar{\omega}, 0) \in \mathbb{R}^d \times \mathbb{R}^{n-d}$ with $\bar{\omega}$ non-resonant. For $Q \geq n+2$ and $\sqrt{\varepsilon} \leq \rho \leq 1$, assume that*

$$(116) \qquad \rho Q\Psi(Q)s^{-1}\cdot < 1, \quad \lambda sQ \cdot > 1,$$

for some $\lambda \cdot = 1$. Then there exists a $(M, s/4)$-ultra-differentiable symplectic transformation

$$\Phi : \mathbb{T}^n \times D_{\rho/2} \to \mathbb{T}^n \times D_\rho$$

such that

$$H \circ \Phi = h + \bar{f} + \hat{f}$$

where $\{\bar{f}, L_\omega\} = 0$ and with the estimates

$$|\sigma^{-1} \circ \Phi \circ \sigma - \text{Id}|_{s/2} \lessdot \Psi(Q)\varepsilon, \quad |\bar{f} \circ \sigma|_{s/2} \lessdot \varepsilon, \quad |\hat{f} \circ \sigma|_{s/2} \lessdot \varepsilon \exp\left(-\kappa_{2^{1/d}}\left(C^{-1}(\lambda sQ)\right)^{-1}\right).$$

PROOF OF THEOREM J. It suffices to choose

$$Q := \Delta^{-1}\left(c_2 s \rho^{-1}\right), \quad c_2 =\cdot 1$$

and to verify that choosing ρ sufficiently small allows indeed to apply Proposition 31. $\qquad\square$

PROOF OF PROPOSITION 32. To analyze our Hamiltonian H in the domain $\mathbb{T}^n \times D_\rho$, which is a neighborhood of size ρ around the origin in action space, we rescale the action variables using the map

$$\sigma = \sigma_\rho : (\theta, I) \longmapsto (\theta, \rho I)$$

which sends the domain $\mathbb{T}^n \times D$ onto $\mathbb{T}^n \times D_\rho$, the latter being included in $\mathbb{T}^n \times D$ since $\rho \leq 1$. Let

$$H' = \rho^{-1}(H \circ \sigma)$$

be the rescaled Hamiltonian, so H' is defined on $\mathbb{T}^n \times D$ and reads

$$H'(\theta, I) = \rho^{-1} H(\theta, \rho I) = \rho^{-1} h(\rho I) + \rho^{-1} f(\theta, \rho I), \quad (\theta, I) \in \mathbb{T}^n \times D.$$

Without loss of generality, we assume that $h(0) = 0$. Now, using Taylor's formula, we can expand h around the origin to obtain

$$h(\rho I) = \rho \omega \cdot I + \rho^2 \int_0^1 (1-t)\nabla^2 h(t\rho I)(I, I)dt = \rho \omega \cdot I + \rho^2 h'(I)$$

and so we can write

$$\begin{cases} H' = L_\omega + R + F, \\ R := \rho h', \quad F := \rho^{-1}(f \circ \sigma). \end{cases}$$

Now we know that $|f|_s \leq \varepsilon$, so that

$$|\rho^{-1}(f \circ \sigma)|_s \leq \rho^{-1}\varepsilon.$$

Moreover, applying Corollary 9 we have the estimate

$$|\nabla h'|_{s/2} \lessdot s^{-1}|h|_{s/2} \lessdot s^{-1}$$

so eventually our Hamiltonian H' reads

$$\begin{cases} H' = L_\omega + R + F, \\ R := \rho h', \quad F := \rho^{-1}(f \circ \sigma) \\ |R|_{s/2} \lessdot s^{-1}\rho, \quad |F|_{s/2} \leq \rho^{-1}\varepsilon. \end{cases}$$

It has exactly the form (108) with $s/2$ instead of s, a constant times $s^{-1}\rho$ instead of ρ and $\rho^{-1}\varepsilon$ instead of ν; using the assumption that $\sqrt{\varepsilon} \leq \rho$, we have

$$\nu = \rho^{-1}\varepsilon \leq \rho$$

and (116) implies (109). It follows that Proposition 31 can be applied, and it gives the existence of a $(M, s/4)$-ultra-differentiable symplectic transformation

$$\Phi' : \mathbb{T}^n \times D_{1/2} \to \mathbb{T}^n \times D$$

such that

$$H' \circ \Phi' = L_\omega + R + \bar{F} + \hat{F}$$

where $\{\bar{F}, L_\omega\} = 0$ and with the estimates

(117) $|\Phi' - \mathrm{Id}|_{s/4} \lessdot \Psi(Q)\rho^{-1}\varepsilon, \quad |\bar{F}|_{s/4} \lessdot \rho^{-1}\varepsilon, \quad |\hat{F}|_{s/4} \lessdot \rho^{-1}\varepsilon \exp(-\kappa_{2^{1/d}}\left(C^{-1}(\lambda s Q)\right)^{-1}).$

Now, scaling back to our original coordinates, we define

$$\Phi = \sigma \circ \Phi' \circ \sigma^{-1} : \mathbb{T}^n \times D_{\rho/2} \to \mathbb{T}^n \times D_\rho$$

so that

$$\begin{aligned} H \circ \Phi &= \rho H' \circ \Phi' \circ \sigma^{-1} \\ &= \rho(L_\omega + R + \bar{F} + \hat{F}) \circ \sigma^{-1} \\ &= (\rho L_\omega + \rho^2 h') \circ \sigma^{-1} + \rho \bar{F} \circ \sigma^{-1} + \rho \hat{F} \circ \sigma^{-1} \end{aligned}$$

where we used the definition of R. Observe that $(\rho L_\omega + \rho^2 h') \circ \sigma^{-1} = h$, so we may set

$$\bar{f} = \rho \bar{F} \circ \sigma^{-1}, \quad \hat{f} = \rho \hat{F} \circ \sigma^{-1},$$

and write

$$H \circ \Phi = h + \bar{f} + \hat{f}.$$

Since $\{\bar{F}, L_\omega\} = 0$, $\{\bar{F} \circ \sigma^{-1}, L_\omega\} = 0$ and thus $\{\bar{f}, L_\omega\} = 0$. To conclude, observe that the estimates (117) can be written again as

$$\begin{cases} |\sigma^{-1} \circ \Phi \circ \sigma - \mathrm{Id}|_{s/4} \lessdot \Psi(Q)\rho^{-1}\varepsilon, & |\bar{f} \circ \sigma|_{s/4} \lessdot \varepsilon, \\ |\hat{f} \circ \sigma|_{s/4} \lessdot \varepsilon \exp\left(-\kappa_{2^{1/d}} \left(C^{-1}(\lambda s Q)\right)^{-1}\right) \end{cases}$$

which is exactly what we wanted to prove. □

Let us conclude by giving the proof of Corollary 3, which is just a slight variation on the proof of Corollary 1.

PROOF OF COROLLARY 3. Since we are under the assumptions of Theorem J, let us first consider the Hamiltonian in normal form

$$H \circ \Phi = h + \bar{f} + \hat{f}$$

where

$$\Phi : \mathbb{T}^n \times D_{\rho/2} \to \mathbb{T}^n \times D_\rho,$$

$\{\bar{f}, L_\omega\} = 0$ and with the estimates

$$\begin{cases} |\sigma^{-1} \circ \Phi \circ \sigma - \mathrm{Id}|_{s/4} \leq c_1 \Psi(\Delta^{-1}(c_2 s \varepsilon^{-1}))\rho^{-1}\varepsilon, & |\bar{f} \circ \sigma|_{s/4} \leq c_1\varepsilon, \\ |\hat{f} \circ \sigma|_{s/4} \leq c_1\varepsilon \exp(-c_3 \left(C^{-1}(c_4 s \Delta^{-1}(c_2 s \rho^{-1}))\right)^{-1}). \end{cases}$$

Decomposing $\Phi = (\Phi_\theta, \Phi_I)$ into its angle and action space components, and observing that $\partial_\theta(\hat{f} \circ \sigma) = (\partial_\theta \hat{f}) \circ \sigma$, one obtains from these last estimates and Corollary 9 that

$$|\Phi_I - \mathrm{Id}_I|_{s/4} \leq c_1 \Psi(\Delta^{-1}(c_2 s \varepsilon^{-1}))\varepsilon$$

and

$$\sup_{(\tilde{\theta}, \tilde{I}) \in \mathbb{T}^n \times D_{\rho/2}} |\partial_\theta \hat{f}(\tilde{\theta}, \tilde{I})| \lessdot s^{-1}\varepsilon \exp(-c_3 \left(C^{-1}(c_4 s \Delta^{-1}(c_2 s \rho^{-1}))\right)^{-1}).$$

Let $(\tilde{\theta}(t), \tilde{I}(t))$ be a solution of $H \circ \Phi$ with $\tilde{I}(0) \in D_{\rho/4}$. Since $\{\bar{f}, L_\omega\} = 0$ and $\omega = (\bar{\omega}, 0) \in \mathbb{R}^d \times \mathbb{R}^{n-d}$ with $\bar{\omega}$ non-resonant, we have $\partial_{\theta_j}\bar{f} = 0$ for $1 \leq j \leq d$ and thus

$$|\Pi_d(\tilde{I}(t) - \tilde{I}(0))| \leq |t| \sup_{(\tilde{\theta}, \tilde{I}) \in \mathbb{T}^n \times D_{\rho/2}} |\partial_\theta \hat{f}(\tilde{\theta}, \tilde{I})|$$

for all t as long as $\tilde{I}(t) \in D_{\rho/4}$. So if we define

$$T := rs\varepsilon^{-1} \exp(c_3 \left(C^{-1}(c_4 s \Delta^{-1}(c_2 s \varepsilon^{-1}))\right)^{-1})$$

we obtain

$$|\Pi_d(\tilde{I}(t) - \tilde{I}(0))| \leq r, \quad |t| \leq T$$

since $r \leq \rho/4$ as long as $\tilde{I}(t) \in D_{\rho/2}$. Now coming back to the original Hamiltonian H, any solution $(\theta(t), I(t))$ with $I(0) \in D_{\rho/8}$ gives rise to a solution $(\tilde{\theta}(t), \tilde{I}(t)) = \Phi^{-1}(\theta(t), I(t))$ of $H \circ \Phi$ with $\tilde{I}(0) \in D_{\rho/4}$, and therefore, for $|t| \leq T$ and such that $I(t) \in D_{\rho/4}$, we obtain

$$\begin{aligned} |\Pi_d(I(t) - I(0))| &\leq |\Pi_d(I(t) - \tilde{I}(t))| + |\Pi_d(\tilde{I}(t) - \tilde{I}(0))| + |\Pi_d(\tilde{I}(0) - I(0))| \\ &\leq r + 2c_1 \Psi(\Delta^{-1}(c_2 s \varepsilon^{-1}))\varepsilon \\ &\leq 2r \end{aligned}$$

since we assumed $r \geq 2c_1 \Psi(\Delta^{-1}(c_2 s \varepsilon^{-1}))\varepsilon$. This concludes the proof. □

5. Stability in the quasi-convex case: proof of Theorem K

The aim of this section is to prove Theorem K; implicit constants will depend only on n, $|h|_s$ and the function C (except at the very end of the proof where they will also depend on l and m).

To prove Theorem K, we will have to establish several intermediate statements. The first one is a normal form for an arbitrary non-linear Hamiltonian in a neighborhood of a periodic frequency. In order not to assume that h is a local diffeomorphism, the following definition will be useful in the sequel. For a fixed integrable Hamiltonian h and for $\mu > 0$, a point $I_1 \in D$ in action space is said to be μ-close to be T-periodic if there exists a T-periodic vector $v \in \mathbb{R}^n \setminus \{0\}$ such that

$$(118) \qquad |\nabla h(I_1) - v| \le \mu.$$

We will use the notation $D_r(I_1)$ to denote the ball of radius $r > 0$ centered at the point I_1. We have the following proposition, where we define

$$\tau_1(\theta, I) := (\theta, I - I_1)$$

and we recall that, given $\rho > 0$,

$$\sigma(\theta, I) = \sigma_\rho(\theta, I) = (\theta, \rho I).$$

PROPOSITION 33. *Let H be as in $(*)$, where H is (M, s)-ultra-differentiable with M satisfying (H1) and (H2). Let $I_1 \in D_{3/4}$ be μ-close to be T-periodic, and given $\rho > 0$, assume that*

$$(119) \qquad \mu s \le \rho, \quad \varepsilon \cdot < \rho^2, \quad T\rho \cdot < s^2, \quad \rho < 1/4.$$

Then there exists a $(M, s/4)$-ultra-differentiable symplectic transformation

$$\Phi : \mathbb{T}^n \times D_{\rho/2}(I_1) \to \mathbb{T}^n \times D_\rho(I_1)$$

such that

$$H \circ \Phi = h + \bar{f} + \hat{f}$$

where $\{\bar{f}, L_v\} = 0$ and with the estimates

$$\begin{cases} |\tau_1^{-1} \circ \sigma^{-1} \circ \Phi \circ \sigma \circ \tau_1 - \mathrm{Id}|_{s/4} \le 2T\rho^{-1}\varepsilon, \quad |\bar{f} \circ \sigma \circ \tau_1|_{s/4} \le 2\varepsilon, \\ |\hat{f} \circ \sigma \circ \tau_1|_{s/4} \le \varepsilon \exp\left(-\kappa_2 \left(C^{-1}(s^2(AT\rho)^{-1})\right)^{-1}\right). \end{cases}$$

PROOF. Observe that since $\rho < 1/4$ and $I_1 \in D_{3/4}$, then $D_\rho(I_1)$ is included in D. Without loss of generality, we may assume that $I_1 = 0$ so that τ_1 is the identity, and we may also assume that $h(0) = 0$. Let

$$H' = \rho^{-1}(H \circ \sigma)$$

be the rescaled Hamiltonian, so H' is defined on $\mathbb{T}^n \times D$ and reads

$$H'(\theta, I) = \rho^{-1} H(\theta, \rho I) = \rho^{-1} h(\rho I) + \rho^{-1} f(\theta, \rho I), \quad (\theta, I) \in \mathbb{T}^n \times D.$$

Using Taylor's formula, we can expand h around the origin to obtain

$$h(\rho I) = \rho\omega \cdot I + \rho^2 \int_0^1 (1-t)\nabla^2 h(t\rho I)(I, I) dt = \rho\omega \cdot I + \rho^2 h'(I)$$

which can be written again as

$$h(\rho I) = \rho v \cdot I + \rho(\omega - v) \cdot I + \rho^2 h'(I)$$

and therefore

$$\begin{cases} H' = L_v + S + R + F, \\ S := L_\omega - L_v, \quad R := \rho h', \quad F := \rho^{-1}(f \circ \sigma). \end{cases}$$

From the assumptions that I_1 is μ-close to be T-periodic we know that

$$|\nabla S|_s \le c\mu.$$

Then as $|f|_s \le \varepsilon$, we have

$$|\rho^{-1}(f \circ \sigma)|_s \le \rho^{-1}\varepsilon$$

and moreover, applying Corollary 9, we have the estimate

$$|\nabla h'|_{s/2} \lessdot s^{-1}.$$

Eventually our Hamiltonian H' reads

$$\begin{cases} H' = L_\omega + S + R + F, \\ R := \rho h', \quad S := L_\omega - L_v, \quad F := \rho^{-1}(f \circ \sigma) \\ |\nabla S|_{s/2} \le c\mu, \quad |R|_{s/2} \lessdot s^{-1}\rho, \quad |F|_{s/2} \le \rho^{-1}\varepsilon. \end{cases}$$

It has exactly the form ($\sharp$) with $s/2$ instead of s, $c\mu$ instead of μ, a constant times $s^{-1}\rho$ instead of ρ and $\rho^{-1}\varepsilon$ instead of ν; one easily check that (119) implies $\eta =\cdot \max\{\mu, s^{-1}\rho\} =\cdot s^{-1}\rho$ and (97). It follows that Proposition 27 can be applied, and it gives, choosing $\delta = 1$ and $\xi = 2$, the existence of a $(M, s/4)$-ultra-differentiable symplectic transformation

$$\Phi' : \mathbb{T}^n \times D_{1/2} \to \mathbb{T}^n \times D$$

such that

$$H \circ \Phi' = L_v + S + \bar{F} + \hat{F}$$

with $\{\bar{F}, L_v\} = 0$ and the estimates

(120)
$$\begin{cases} |\Phi' - \mathrm{Id}|_{s/4} \le 2T\rho^{-1}\varepsilon, \quad |\bar{F}|_{s/4} \le 2\rho^{-1}\varepsilon, \\ |\hat{F}|_{s/4} \le \rho^{-1}\varepsilon \exp\left(-\kappa_2 \left(C^{-1}(s^2(AT\rho)^{-1})\right)^{-1}\right). \end{cases}$$

Now, scaling back to our original coordinates, we define

$$\Phi = \sigma \circ \Phi' \circ \sigma^{-1} : \mathbb{T}^n \times D_{\rho/2} \to \mathbb{T}^n \times D_\rho$$

so that

$$\begin{aligned} H \circ \Phi &= \rho H' \circ \Phi' \circ \sigma^{-1} \\ &= \rho(L_\omega + R + \bar{F} + \hat{F}) \circ \sigma^{-1} \\ &= (\rho L_\omega + \rho^2 h') \circ \sigma^{-1} + \rho\bar{F} \circ \sigma^{-1} + \rho\hat{F} \circ \sigma^{-1} \end{aligned}$$

where we used the definition of R. Observe that $(\rho L_\omega + \rho^2 h') \circ \sigma^{-1} = h$, so we may set

$$\bar{f} = \rho\bar{F} \circ \sigma^{-1}, \quad \hat{f} = \rho\hat{F} \circ \sigma^{-1},$$

and write

$$H \circ \Phi = h + \bar{f} + \hat{f}.$$

Since $\{\bar{F}, L_\omega\} = 0$, $\{\bar{F} \circ \sigma^{-1}, L_\omega\} = 0$ and thus $\{\bar{f}, L_\omega\} = 0$. To conclude, observe that the estimates (120) can be written again as

$$\begin{cases} |\sigma^{-1} \circ \Phi \circ \sigma - \mathrm{Id}|_{s/4} \le 2T\rho^{-1}\varepsilon, \quad |\bar{f} \circ \sigma|_{s/4} \le 2\varepsilon, \\ |\hat{f} \circ \sigma|_{s/4} \le \varepsilon \exp\left(-\kappa_2 \left(C^{-1}(s^2(AT\rho)^{-1})\right)^{-1}\right) \end{cases}$$

which is exactly what we wanted to prove. $\square$

From the last proposition, it is easy to prove that the action variables of any solution starting close to I_1 have small variation for a long interval of time in the direction given by the periodic vector v. Using the quasi-convexity assumption on h, we will prove that the action variables have also small variation for a long interval of time in the directions transverse to v but tangent to the energy level. Let us recall that $h : D \to \mathbb{R}$ is (l, m)-quasi-convex, for positive constants l and m, if for all $I \in D$, it satisfies

$$|\nabla h(I)| \ge l$$

and at least of the inequalities

$$|\nabla h(I) \cdot \xi| \ge l|\xi|, \quad |\nabla^2 h(I)\xi \cdot \xi| \ge m|\xi|^2$$

holds true for all $\xi \in \mathbb{R}^n$.

PROPOSITION 34. *Under the assumptions of Proposition 33 and if*

(121)
$$\mu \cdot< m\rho, \quad \rho \cdot< l, \quad \varepsilon \cdot< m\rho^2,$$

for any $I_0 \in D_\mu(I_1)$ and any solution $(\theta(t), I(t))$ of H with $I(0) = I_0$, we have

$$|I(t) - I_0| \leq \rho, \quad |t| \cdot< m\rho^2 s\varepsilon^{-1} \exp(\kappa_2 \left(C^{-1}(s^2(AT\rho)^{-1})\right)^{-1})$$

where $A = \cdot 1$.

The proof below follows [**Pös93**].

PROOF. Let us define

$$T_* \cdot= m\rho^2 s\varepsilon^{-1} \exp(\kappa_2 \left(C^{-1}(s(AT\rho)^{-1})\right)^{-1}).$$

For any solution $(\tilde{\theta}(t), \tilde{I}(t))$ of $H \circ \Phi$ with $\tilde{I}(0) \in D_{2\mu}(I_1)$, let $T_e \in]0, +\infty]$ be the time of first exit of the ball centered at I_0 of radius $\rho/4$; observe that $\mu \leq \rho/8$ can be arranged by the first inequality in (121) so this is well-defined. We claim it is sufficient to prove that $T_* \leq T_e$. Indeed, in this case one would have

$$|\tilde{I}(t) - \tilde{I}(0)| \leq \rho/4, \quad 0 \leq t \leq T_*$$

and as the image of $D_{2\mu}(I_1)$ (respectively $D_{\rho/2}(I_1)$) under Φ contains $D_\mu(I_1)$ (respectively is contained in $D_\rho(I_1)$), one would obtain that

$$|I(t) - I(0)| \leq \rho, \quad 0 \leq t \leq T_*$$

for any solution $(\theta(t), I(t))$ of H with $I(0) \in D_\mu(I_1)$.

So it remains to prove the claim, and we will do this by contradiction assuming that $T := \min\{T_e, T_*\} = T_e$. To simplify the notations, we remove the tilde from the notations and we consider a solution $(\theta(t), I(t))$ of $H \circ \Phi$ with $I(0) \in D_{2\mu}(I_1)$. Let us write

$$\Delta I := I(T) - I(0), \quad I(s) := I_0 + s\Delta(I), \quad \omega(I(s)) := \nabla h(I(s))$$

for $0 \leq s \leq 1$. Observe that since we are assuming $T = T_e$, then

(122)
$$|\Delta I| = \rho/4.$$

Recalling that

$$|\nabla h(I_1) - v| \leq \mu$$

for some T-periodic v, we let Π_v be the orthogonal projection onto the line generated by v and $\Pi_v^\perp = \mathrm{Id} - \Pi_v$ be the orthogonal projection onto the orthogonal of v. Since $I(0) \in D_{2\mu}(I_1)$, we have

(123)
$$|\omega(I(0)) - \nabla h(I_1)| \cdot< \mu$$

and using (122) and the fact that $\mu \leq \rho$, we obtain

$$|\omega(I(s)) - \nabla h(I_1)| \cdot< \rho.$$

So from the last two inequalities

(124)
$$|\omega(I(s)) - v| \cdot< \rho.$$

Since $\Pi_v^\perp v = 0$ we have

$$|\Pi_v^\perp \omega(I(s))| = |\Pi_v^\perp(\omega(I(s) - v))| \cdot< \rho$$

and so

(125)
$$|\omega(I(s)) \cdot \Pi_v^\perp \Delta I| \cdot< |\Pi_v^\perp(\omega(I(s) - v))||\Pi_v^\perp \Delta I| \cdot< \rho|\Delta I|.$$

Observe that for $s = 0$, we have the better estimate

(126)
$$|\omega(I(0)) \cdot \Pi_v^\perp \Delta I| \cdot< \mu|\Delta I|.$$

On the other hand, our Hamiltonian is of the form

$$H \circ \Phi = h + \bar{f} + \hat{f}$$

with $\{\bar{f}, L_v\} = 0$, which implies that $\Pi_v(\partial_\theta \bar{f}) = 0$ and thus

$$|\omega(I(s)) \cdot \Pi_v \Delta I| \leq \int_0^T |\omega(I(s)) \cdot \partial_\theta \hat{f}| dt.$$

Using Corollary 9 one easily obtains an estimate of $\partial_\theta \hat{f}$ from the known estimate on $\hat{f}$, and, by the definition of T with a proper implicit constant, we arrive at

$$(127) \qquad |\omega(I(s)) \cdot \Pi_v \Delta I| < m\rho^2/96 = (m\rho/24)\,|\Delta I|$$

where the last equality follows from (122). From (125) and (127) we get

$$|\omega(I(s)) \cdot \Delta I| \lessdot \rho |\Delta I|$$

and by a proper choice of the implicit constant in the second condition of (121), we can ensure that

$$(128) \qquad |\omega(I(s)) \cdot \Delta I| \leq l|\Delta I|.$$

To reach a contradiction, let $\Delta h := h(I(T)) - h(I(0))$ so that Taylor formula with integral remainder at the point $I(0)$ gives

$$\Delta h = \omega(I(0)) \cdot \Delta I + \int_0^1 (1-s)\nabla^2 h(I(s))\Delta I \cdot \Delta I ds.$$

Since h is (l, m)-convex, using (128) we know that

$$|\nabla^2 h(I(s))\Delta I \cdot \Delta I| \geq m|\Delta I|^2$$

and therefore

$$(129) \qquad m|\Delta I|^2 \leq 2|\Delta h| + 2|\omega(I(0)) \cdot \Delta I| \leq 2|\Delta h| + 2|\omega(I(0)) \cdot \Pi_v^\perp \Delta I| + 2|\omega(I(0)) \cdot \Pi_v \Delta I|.$$

Using the preservation of energy, the last part of (121) and (122), we have

$$(130) \qquad 2|\Delta h| \lessdot \varepsilon < \frac{m}{3}\frac{\rho^2}{16} = \frac{m}{3}|\Delta I|^2.$$

Next, from (126) we obtain

$$2|\omega(I(0)) \cdot \Pi_v^\perp \Delta I| \lessdot \mu|\Delta I| \leq C\frac{\mu^2}{m} + \frac{m}{6}|\Delta I|^2$$

for some $C = \cdot 1$, and, using the first part of (121) and (122), we can ensure that

$$(131) \qquad 2|\omega(I(0)) \cdot \Pi_v^\perp \Delta I| < \frac{m}{6}|\Delta I|^2 + \frac{m}{6}|\Delta I|^2 = \frac{m}{3}|\Delta I|^2.$$

Finally, recall from (127) that

$$(132) \qquad 2|\omega(I(s)) \cdot \Pi_v \Delta I| < m\rho^2/48 = \frac{m}{3}|\Delta I|^2.$$

and thus (129), (130), (131) and (132) yields

$$m|\Delta I|^2 < m|\Delta I|^2$$

which is absurd.

This proves the proposition for positive times, but for negative times the argument is completely similar so this concludes the proof. $\square$

To finish the proof of Theorem K, we just need the following lemma which is a direct application of Dirichlet's box principle.

LEMMA 35. *Let $\omega \in \mathbb{R}^n \setminus \{0\}$, and $Q \geq 1$ a real number. Then there exists a T-periodic vector $v \in \mathbb{R}^n \setminus \{0\}$ such that*

$$|\omega - v| \leq (n-1)(TQ)^{-1}, \quad |\omega|^{-1} \leq T \leq n|\omega|^{-1}Q^{n-1}.$$

PROOF. Fix $Q \geq 1$. Up to a re-ordering of its component, we can write $\omega = |\omega|_\infty(\pm 1, x)$ for some $x \in \mathbb{R}^{n-1}$ and by Dirichlet's approximation theorem, there exists a primitive rational vector $p/q \in \mathbb{Q}^{n-1}$, such that

$$|qx - p|_\infty \leq Q^{-1}, \quad 1 \leq q \leq Q^{n-1}.$$

The vector $v = |\omega|_\infty(\pm 1, p/q) \in \mathbb{R}^n$ is then T-periodic, for $T = |\omega|_\infty^{-1} q$, and we have

$$|\omega - v| \leq T^{-1}|qx - p|, \quad |\omega|_\infty^{-1} \leq T \leq |\omega|_\infty^{-1} Q^{n-1}$$

which implies

$$|\omega - v| \leq (n-1)(TQ)^{-1}, \quad |\omega|^{-1} \leq T \leq n|\omega|^{-1}Q^{n-1}$$

and this was the statement to prove.   □

PROOF OF THEOREM K. Let $Q \geq 1$ be a real number to be chosen below. Recall that we are considering H as in $(*)$ where H is (M, s)-ultra-differentiable with M satisfying (H1) and (H2) and h is (m, l)-quasi-convex. To further simplify the exposition, implicit constants will now also depend on l and m.

Consider any solution $(\theta(t), I(t))$ of H with $I(0) \in D_{1/2}$. We can apply Lemma 35 to $\omega := \nabla h(I(0))$ and find a T-periodic vector $v \in \mathbb{R}^n \setminus \{0\}$ such that

$$|\omega - v| \lessdot (TQ)^{-1}, \quad 1 \lessdot T \lessdot Q^{n-1}.$$

We set

$$\mu = \cdot (TQ)^{-1}$$

so that $I(0)$ is μ-close to be T-periodic. Then we set

$$\rho = \cdot s\mu = \cdot s(TQ)^{-1}$$

and using the fact that $T \lessdot Q^{n-1}$, (119) and (121) are satisfied provided that

(133) $Q^{2n}\varepsilon \lessdot s^2, \quad Qs \cdot> 1.$

Assuming (133), Proposition 33 and Proposition 34 apply, and we obtain the stability estimates

$$|I(t) - I(0)| \leq \rho \lessdot sQ^{-1},$$

for

$$|t| \lessdot s \exp\left(\kappa_2 \left(C^{-1}(aQs)\right)^{-1}\right), \quad a \cdot= 1.$$

To conclude, to fulfill (133) we may choose

$$Q \cdot= (s^2\varepsilon^{-1})^{\frac{1}{2n}}$$

to finally reach

$$|I(t) - I(0)| \leq \rho \lessdot s(s^{-2}\varepsilon)^{\frac{1}{2n}},$$

for

$$|t| \lessdot s \exp\left(\kappa_2 \left(C^{-1}\left(a's(s^2\varepsilon^{-1})^{\frac{1}{2n}}\right)\right)^{-1}\right), \quad a' \cdot= 1$$

which was the statement to prove.   □

6. Diffusion in the quasi-convex case: proof of Theorem L

The goal of this section is to give a proof of Theorem L, which is the construction of an unstable orbit assuming that we are working in a non-quasi analytic class of ultra-differentiable functions (which are characterized by the condition (H3), see § 6).

The construction, which is due to Marco-Sauzin, is explained in details in [**MS02**] in the case where $M = M_\alpha$ with $\alpha > 1$; we are aiming at showing that it still works in any non-quasi-analytic class of ultra-differentiable functions, and it requires only a minor modification (in the same way as Theorem B required a minor modification of [**Bes00**]).

The interest is that when the sequence M is matching, the instability result is very close to the stability result given by Theorem K, and thus this shows that the latter result is quite accurate.

From now on, we write $\mathbb{A} := \mathbb{T} \times \mathbb{R}$ the infinite annulus. Given a potential function $U : \mathbb{T} \to \mathbb{R}$, we consider the following family of maps $\psi_q : \mathbb{A} \to \mathbb{A}$ defined by

$$(134) \qquad \psi_q(\theta, I) = \left(\theta + qI, I - q^{-1}U'(\theta + qI)\right), \quad (\theta, I) \in \mathbb{A},$$

for $q \in \mathbb{N}^*$. If we require $U'(0) = -1$, for example if we choose

$$U(\theta) = -(2\pi)^{-1}\sin(2\pi\theta), \quad U'(\theta) = -\cos(2\pi\theta),$$

then it is easy to see that $\psi_q(0,0) = (0, q^{-1})$ and by induction

$$(135) \qquad \psi_q^k(0,0) = (0, kq^{-1})$$

for any $k \in \mathbb{Z}$. After q iterations, the point $(0,0)$ drifts from the circle $I = 0$ to the circle $I = 1$ and it is bi-asymptotic to infinity, in the sense that the sequence $\left(\psi_q^k(0,0)\right)_{k \in \mathbb{Z}}$ is not contained in any semi-infinite annulus of $\mathbb{A}$. Clearly these maps are exact-symplectic, but obviously they have no invariant circles and so they cannot be "close to integrable". However, we will use the fact that they can be written as a composition of time-one maps,

$$(136) \qquad \psi_q = \Phi^{q^{-1}U} \circ \left(\Phi^{\frac{1}{2}I^2} \circ \cdots \circ \Phi^{\frac{1}{2}I^2}\right) = \Phi^{q^{-1}U} \circ \left(\Phi^{\frac{1}{2}I^2}\right)^q,$$

to embed ψ_q in the qth-iterate of a near-integrable map of $\mathbb{A}^n$, for $n \geq 2$. To do so, we will use the following "coupling lemma", in which the product of functions acting on separate variables is denoted by $\otimes$, $i.e.$

$$f \otimes g(x, x') = f(x)g(x') \quad x \in \mathbb{A}^m, \; x' \in \mathbb{A}^{m'}.$$

LEMMA 36 (Marco-Sauzin). *Let $m, m' \geq 1$, $F : \mathbb{A}^m \to \mathbb{A}^m$ and $G : \mathbb{A}^{m'} \to \mathbb{A}^{m'}$ two maps, and $f : \mathbb{A}^m \to \mathbb{R}$ and $g : \mathbb{A}^{m'} \to \mathbb{R}$ two Hamiltonian functions generating complete vector fields. Suppose there is a point $a \in \mathbb{A}^{m'}$ which is q-periodic for G and such that the following "synchronisation" conditions hold:*

$$(S) \qquad g(a) = 1, \quad dg(a) = 0, \quad g(G^k(a)) = 0, \quad dg(G^k(a)) = 0,$$

for $1 \leq k \leq q - 1$. Then the mapping

$$\Psi = \Phi^{f \otimes g} \circ (F \times G) : \mathbb{A}^{m+m'} \longrightarrow \mathbb{A}^{m+m'}$$

is well-defined and for all $x \in \mathbb{A}^m$,

$$\Psi^q(x, a) = (\Phi^f \circ F^q(x), a).$$

Therefore, if we set $m = 1$, $F = \Phi^{\frac{1}{2}I_1^2}$ and $f = q^{-1}U$ in the coupling lemma, the qth-iterate Ψ^q will leave the submanifold $\mathbb{A} \times \{a\}$ invariant, and its restriction to this annulus will coincide with our "unstable map" ψ_q. Hence, after q^2 iterations of Ψ, the I_1-component of the point $((0,0), a) \in \mathbb{A}^2$ will move from 0 to 1.

The difficult part is then to find what kind of dynamics we can put on the second factor to apply this coupling lemma. Following [**MS02**], we introduce a new sequence of "large" parameters $A_j \in \mathbb{N}^*$ for $j \in \mathbb{N}$ and in the second factor we consider a family of suitably rescaled penduli on $\mathbb{A}$ given by

$$P_j(\theta_2, I_2) = \frac{1}{2}I_2^2 + A_j^{-2}V(\theta_2),$$

where $V(\theta) = -\cos 2\pi\theta$. We eventually define

$$G_j = \Phi^{\frac{1}{2}(I_2^2 + I_3^2 + \cdots + I_n^2) + N_j^{-2}V(\theta_2)}.$$

In this case, the map Ψ given by the coupling lemma is a perturbation of

$$\Phi^h = \Phi^{\frac{1}{2}(I_1^2 + I_2^2 + \cdots + I_n^2)}$$

of the form $\Psi = \Phi^u \circ \Psi^{\tilde{h}+v}$, with

$$u = f \otimes g, \quad v = A_j^{-2}V.$$

But for the map G_j, due to the presence of the pendulum factor, it is possible to find a periodic orbit for G_j with an irregular distribution: more precisely, a q_j-periodic point a_j such that its distance to the rest of its orbit is of order N_j^{-1}, no matter how large q_j is.

Let us denote by $(p_j)_{j \geq 0}$ the ordered sequence of prime numbers and let us choose $A_j = p_j A'_j$ where $A'_j = 1$ if $n = 2$ and

$$A'_j = p_{j-(n-3)} p_{j-(n-4)} \cdots p_j \in \mathbb{N}^*, \quad n \geq 3.$$

Our goal is to prove the following proposition.

PROPOSITION 37. *Let $n \geq 2$ and $s > 0$. Then for all $j \in \mathbb{N}$, there exist a sequence of functions $g_j \in \mathcal{U}_s(\mathbb{A}^{n-1})$, a sequence of points $a_j \in \mathbb{T}^{n-1}$ and positive constants c_1 and s' depending only on s and M such that if*

$$B_j = 2\left[c_1 c^{2(n-1)} A_j \exp(2\Omega(s' p_j)) + 1 \right], \quad q_j = A_j B_j,$$

then a_j is q_j-periodic for G_j and (g_j, G_j, a_j, q_j) satisfy the synchronization conditions (S):

$$g_j(a_j) = 1, \quad dg_j(a_j) = 0, \quad g_j(G_j^k(a_j)) = 0, \quad dg_j(G_j^k(a_j)) = 0,$$

for $1 \leq k \leq q_j - 1$. Moreover, the estimate

(137) $$q_j^{-1} |g_j|_s \leq A_j^{-2},$$

holds true.

We claim that to prove Theorem L, it is enough to justify Proposition 37. Indeed, it is rather easy to obtain a discrete version of Theorem L for maps on $\mathbb{A}^n$, using the prime number Theorem which ensures that

$$A_j^{\frac{1}{n-1}} \leq p_j \leq 2 A_j^{\frac{1}{n-1}}$$

for j large enough, and hence, setting $\varepsilon_j = A_j^{-2}$

$$\varepsilon_j^{-\frac{1}{2(n-1)}} \leq p_j \leq 2 \varepsilon_j^{-\frac{1}{2(n-1)}}.$$

Then by an elementary suspension argument (using the existence of bump functions for non-quasi-analytic ultra-differentiable functions) one arrives at Theorem L with a quasi-convex integrable Hamiltonian in $n + 1$ degrees of freedom. This is described in details in [**MS02**] for the Gevrey case and works exactly in the same way in our context, so we shall not give more details and concentrate on the proof of Proposition 37.

We first consider the simple pendulum

$$P(\theta, I) = \frac{1}{2} I^2 + V(\theta), \quad (\theta, I) \in \mathbb{A}.$$

With our convention, the stable equilibrium is at $(0, 0)$ and the unstable one is at $(0, 1/2)$. Given any $B \in \mathbb{N}^*$, there is a unique point $b^B = (0, I_B)$ which is B-periodic for Φ^P (this is just the intersection between the vertical line $\{0\} \times \mathbb{R}$ and the closed orbit for the pendulum of period B). One can check that $I_B \in]2, 3[$ and as B goes to infinity, $(0, I_B)$ tends to the point $(0, 2)$ which belongs to the upper separatrix. Since $P_j(\theta, I) = \frac{1}{2} I^2 + A_j^{-2} V(\theta)$, then one can see that

$$\Phi^{P_j} = \sigma_j^{-1} \circ \Phi^{A_j^{-1} P} \circ \sigma_j,$$

where $\sigma_j(\theta, I) = (\theta, A_j I)$ is the rescaling by A_j in the action components. Therefore the point $b_j^B = (0, A_j^{-1} I_B)$ is q_j-periodic for Φ^{P_j}, for $q_j = A_j B$. Let $(\Phi_t^P)_{t \in \mathbb{R}}$ be the flow of the pendulum, and

$$\Phi_t^P(0, I_B) = (\theta_B(t), I_B(t)).$$

The function $\theta_B(t)$ is analytic. The crucial observation is the following simple property of the pendulum

LEMMA 38. *Let $\vartheta = -\frac{1}{2} + \frac{2}{\pi} \arctan e^\pi < \frac{1}{2}$. For any $B \in \mathbb{N}^*$,*

$$\theta_B(t) \notin [-\vartheta, \vartheta],$$

for $t \in [1/2, B - 1/2]$.

Hence no matter how large B is, most of the points of the orbit of $b^B \in \mathbb{A}$ will be outside the set $\{-\vartheta \leq \theta \leq \vartheta\} \times \mathbb{R}$. The construction of a function that vanishes, as well as its first derivative, at these points, will be easily arranged by means of a function, depending only on the angle variables, with support in $\{-\vartheta \leq \theta \leq \vartheta\}$.

As for the other points, it is convenient to introduce the function

$$\tau_B : [-\vartheta, \vartheta] \longrightarrow]-1/2, 1/2[$$

which is the analytic inverse of θ_B. One can give an explicit formula for this map:

$$\tau_B(\theta) = \int_0^\theta \frac{d\varphi}{\sqrt{I_B^2 - 4\sin^2 \pi\varphi}}.$$

In particular, it is analytic and therefore it belongs to $\mathcal{U}_s([-\vartheta, \vartheta])$ for any sequence $M := (M_l)_{l \in \mathbb{N}}$ and any $s > 0$, and one can obtain the following estimate (see Lemma 2.3 in [**MS02**] for a proof).

LEMMA 39. *For any sequence $M := (M_l)_{l \in \mathbb{N}}$ and any $s > 0$,*

$$\Lambda = \Lambda(M, s) = \sup_{B \in \mathbb{N}^*} |\tau_B|_s < +\infty.$$

Under the action of τ_B, the points of the orbit of b^B whose projection onto $\mathbb{T}$ belongs to $\{-\vartheta \leq \theta \leq \vartheta\}$ get equi-distributed, and we can use the following elementary lemma, which is the only place where the regularity of the function actually come into play.

LEMMA 40. *For $p \in \mathbb{N}^*$, the analytic function $\eta_p : \mathbb{T} \to \mathbb{R}$ defined by*

$$\eta_p(\theta) = \left(\frac{1}{p} \sum_{l=0}^{p-1} \cos 2\pi l\theta \right)^2$$

satisfies

$$\eta_p(0) = 1, \quad \eta_p'(0) = 0, \quad \eta_p(k/p) = \eta_p'(k/p) = 0,$$

for $1 \leq k \leq p - 1$, and

$$|\eta_p|_s \leq c^2 \exp(2\Omega(8\pi ps)).$$

PROOF. Let

$$\xi_p(\theta) = \frac{1}{p} \sum_{l=0}^{p-1} \cos 2\pi l\theta.$$

For any $l \in \mathbb{N}$, we have

$$\sup_{\theta \in \mathbb{T}} |\xi_p^{(l)}| \leq (2\pi p)^l$$

hence

$$|\xi_p|_s = c \sup_{l \in \mathbb{N}} \frac{(l+1)^2 s^l (2\pi p)^l}{M_l} \leq c \sup_{l \in \mathbb{N}} \frac{(8\pi ps)^l}{M_l} = c \exp(\Omega(8\pi ps))$$

and therefore

$$|\eta_p|_s \leq |\xi_p|_s^2 \leq c^2 \exp(2\Omega(8\pi ps)).$$

$\square$

We can now conclude the proof of Proposition 37.

PROOF OF PROPOSITION 37. Given $s > 0$ and $\vartheta > 0$, consider a bump function $\varphi_\vartheta \in \mathcal{U}_s(\mathbb{T})$ which vanishes identically outside $]-\vartheta, \vartheta[$ and let $c_1 = |\varphi_\vartheta|_s$. We choose our function $g_j \in \mathcal{U}_s(\mathbb{A}^{n-1})$, depending only on the angle variables, of the form

$$g_j = g_j^{(2)} \otimes \cdots \otimes g_j^{(n)},$$

where

$$g_j^{(2)}(\theta_2) = \eta_{p_j}(\tau_{B_j}(\theta_2))\varphi_{s,\vartheta}(\theta_2),$$

and

$$g_j^{(i)}(\theta_i) = \eta_{p_{j-(n-i)}}(\theta_i), \quad 3 \leq i \leq n.$$

Now we choose our point $a_j = (a_j^{(2)}, \ldots, a_j^{(n)}) \in \mathbb{A}^{n-1}$. We set

$$a_j^{(2)} = b_j^{B_j} = (0, A_j^{-1} I_{B_j}),$$

where

$$B_j = 2 \left[c_1 c^{2(n-1)} A_j \exp(2\Omega(s' p_j)) + 1 \right]$$

and

$$a_j^{(i)} = (0, p_{j-(n-i)}^{-1}), \quad 3 \le i \le n.$$

Using the definition of A_j, it is not hard to check that a_j is q_j-periodic for the map G_j.

Now let us show that the synchronization conditions (S) hold true, that is

$$g_j(a_j) = 1, \quad dg_j(a_j) = 0, \quad g_j(G_j^k(a_j)) = 0, \quad dg_j(G_j^k(a_j)) = 0,$$

for $1 \le k \le q_j - 1$. Since $\varphi_\vartheta(0) = 1$, then

$$g_j(a_j) = g_j^{(2)}(0) \cdots g_j^{(n)}(0) = 1$$

and as $\varphi'_\vartheta(0) = 0$, then

$$dg_j(a_j) = 0.$$

To prove the other conditions, let us write $G_j^k(a_j) = (\theta_k, I_k) \in \mathbb{A}^{n-1}$, for $1 \le k \le q_j - 1$.

If $\theta_k^{(2)}$ does not belong to $]-\vartheta, \vartheta[$, then $g_j^{(2)}$ and its first derivative vanish at $\theta_k^{(2)}$ because it is the case for φ_ϑ, so

$$g_j(\theta_k) = dg_j(\theta_k) = 0.$$

Otherwise, if $-\vartheta < \theta_k^{(2)} < \vartheta$, one can easily check that

$$-\frac{A_j - 1}{2} \le k \le \frac{A_j - 1}{2}$$

and therefore

$$\tau_{B_j}(\theta_k^{(2)}) = \frac{k}{A_j},$$

while

$$\theta_k^{(i)} = \frac{k}{p_{j-(n-i)}}, \quad 3 \le i \le n.$$

If A'_j divides k, that is $k = k' A'_j$ for some $k' \in \mathbb{Z}$, then

$$\tau_{B_j}(\theta_k^{(2)}) = \frac{k}{A_j} = \frac{k'}{p_j}$$

and therefore, by Lemma 40, η_{p_j} vanishes with its differential at $\theta_k^{(2)}$, and so does $g_j^{(2)}$. Otherwise, A'_j does not divide k and then, for $3 \le i \le n$, at least one of the functions $\eta_{p_{j-(n-i)}}$ vanishes with its differential at $\theta_k^{(2)}$, and so does $g_j^{(i)}$. Hence in any case

$$g_j(\theta_k) = dg_j(\theta_k) = 0, \quad 1 \le k \le q_j - 1,$$

and the synchronization conditions (S) are satisfied.

Now it remains to estimate the norm of the function g_j. First, by definition of c_1 and using Proposition 10, one finds

$$|g_j|_s \le c_1 |\eta_{p_j} \circ \tau_{B_j}|_s |\eta_{p_{j-(n-3)}}|_s \cdots |\eta_{p_j}|_s.$$

Then, using Lemma 39, Proposition 16 and Lemma 40 we obtain

$$|\eta_{p_j} \circ \tau_{B_j}|_s \le c^2 \exp(2\Omega(s' p_j))$$

for some s' which depends only on s and Λ. By Lemma 40 for $3 \le i \le n$, since $p_{j-(n-i)} \le p_j$, we have

$$|\eta_{p_{j-(n-i)}}|_s \le c^2 \exp(2\Omega(s' p_j))$$

and therefore

$$|g_j|_s \le c_1 c^{2(n-1)} \exp(2(n-1)\Omega(s' p_j))$$

Finally, by definition of B_j we obtain

$$|g_j|_s \le B_j A_j^{-1},$$

and as $q_j = A_j B_j$, we end up with

$$q_j^{-1}|g_j|_s \le A_j^{-2}.$$

This concludes the proof. $\qquad\square$

7. Stability in the steep case: proof of Theorem M

The goal of this section is to prove Theorem M; implicit constants will depend only on n, $|h|_s$, l, L, δ, p and the function C.

Given positive real parameters s, δ, μ, ν, consider a Hamiltonian of the form

$$(\#) \qquad \begin{cases} H(\theta, I) = L_v(I) + g(I) + f(\theta, I), \\ H : \mathbb{T}^n \times D_\delta \to \mathbb{R} \\ |\nabla g|_s \le \rho, \quad |f|_s \le \nu, \end{cases}$$

where $\{\,.\,\}$ denotes the usual Poisson bracket, $D_\delta \subset \mathbb{R}^n$ is the ball of radius δ around the origin. Observe that this is nothing but a special case of $(\sharp)$, up to a change of notations, and thus the following proposition is just a special case of Proposition 27, written in a slightly different way using a real parameter $K \ge 1$.

PROPOSITION 41. Let H be as in $(\#)$, with M satisfying (H1) and (H2) and v a T-periodic vector. Given $K \ge 1$, assume that

$$(138) \qquad\qquad \nu \cdot< s\rho, \quad T\rho \cdot< s, \quad KT\rho \cdot< 1.$$

Then there exists a $(M, s/2)$-ultra-differentiable symplectic transformation

$$\Psi : \mathbb{T}^n \times D_{\delta/2} \to \mathbb{T}^n \times D_\delta$$

such that

$$H \circ \Psi = L_v + g + \bar{f} + \hat{f}$$

with $\{\bar{f}, L_v\} = 0$ and the estimates

$$(139) \qquad |\Psi - \mathrm{Id}|_{s/2} \le 2T\nu, \quad |\bar{f}|_{s/2} \le 2\nu, \quad |\hat{f}|_{s/2} \le \nu \exp\left(-\kappa_2 \left(C^{-1}(Ks)\right)^{-1}\right).$$

Moreover, if I is an integrable Hamiltonian such that $\{I, F\} = 0$, then $\{I, \bar{F}\} = 0$.

Next consider T_i-periodic vectors v_i for every $1 \le i \le n$, and assume that they are linearly independent. For simplicity, let us just write $L_i(I) = v_i \cdot I$. Now consider the following Hamiltonian:

$$(\#_j) \qquad \begin{cases} H_j(\theta, I) = L_j(I) + g_j(I) + f(\theta, I), \\ H_j : \mathbb{T}^n \times D_{\delta_j} \to \mathbb{R} \\ |\nabla g_j|_{s_j} \le \rho_j, \quad |f|_{s_j} \le \nu \end{cases}$$

where we set $\delta_j = 2^{j-1}$ and $s_j = s\delta_j$.

PROPOSITION 42. Let $1 \le j \le n$, H_j be as in $(\#_j)$ with M satisfying (H1) and (H2). Given $K \ge 1$, assume that

$$(140) \qquad \begin{cases} \nu \cdot< s\rho_i, \quad T_i\rho_i \cdot< s, \quad KT_i\rho_i \cdot< 1 \quad 1 \le i \le j, \\ |v_j - v_{j-1}| \cdot< \rho_{j-1}, \quad \rho_j \cdot< \rho_{j-1}, \quad j \ge 2. \end{cases}$$

Then there exists a $(M, s/2)$-ultra-differentiable symplectic transformation

$$\Psi_j : \mathbb{T}^n \times D_{1/2} \to \mathbb{T}^n \times D_{\delta_j}$$

such that

$$H_j \circ \Psi_j = L_j + g + \tilde{f} + f^+$$

with $\{\tilde{f}, L_i\} = 0$ for $1 \leq i \leq j$ and the estimates

$$(141) \qquad |\Psi_j - \mathrm{Id}|_{s/2} \prec \max_{1 \leq i \leq j}\{T_i \nu\}, \quad |\tilde{f}|_{s/2} \prec \nu, \quad |f^+|_{s/2} \prec \nu \exp\left(-\kappa_2 \left(C^{-1}(Ks)\right)^{-1}\right).$$

Moreover, if I is an integrable Hamiltonian such that $\{I, f\} = 0$, then $\{I, \tilde{f}\} = 0$.

PROOF. For $j = 1$, this is nothing but Proposition 41 (setting $\tilde{f} = \bar{f}$ and $f^+ = \hat{f}$) so we may argue by induction. Thus we assume that the statement is true for some $1 \leq j - 1 \leq n - 1$, and we need to show that it remains true for $2 \leq j \leq n$.

To the Hamiltonian $H_j = L_j + g_j + f$ we can apply Proposition 41 (with $L_j, g_j, s_j, \delta_j, \rho_j$ instead of L_v, g, s, δ, ρ) as (141) implies (139): since $s_j/2 = s_{j-1}$ and $\delta_j/2 = \delta_{j-1}$, we find a (M, s_{j-1})-ultra-differentiable symplectic transformation $\Psi^j : \mathbb{T}^n \times D_{\delta_{j-1}} \to \mathbb{T}^n \times D_{\delta_j}$ such that

$$H_j \circ \Psi^j = L_j + g_j + \bar{f} + \hat{f}$$

with $\{\bar{f}, L_j\} = 0$ and with the estimates

$$|\Psi^j - \mathrm{Id}|_{s_{j-1}} \prec T_j \nu, \quad |\bar{f}|_{s_{j-1}} \prec \nu, \quad |\hat{f}|_{s_{j-1}} \prec \nu \exp\left(-\kappa_2 \left(C^{-1}(Ks)\right)^{-1}\right).$$

Moreover, if I is an integrable Hamiltonian such that $\{f, I\} = 0$, then $\{\bar{f}, I\} = 0$.

Then observe that we can write

$$H_j \circ \Psi^j - \hat{f} = L_j + g_j + \bar{f} = L_{j-1} + g_{j-1} + \bar{f}, \quad g_{j-1} := L_j - L_{j-1} + g_j,$$

where

$$|\bar{f}|_{s_{j-1}} \prec \nu$$

and, using (140),

$$|\nabla g_{j-1}|_{s_{j-1}} \leq |\nabla(L_j - L_{j-1})|_{s_{j-1}} + |\nabla g_j|_{s_{j-1}} \prec \rho_{j-1} + \rho_j \prec \rho_{j-1}.$$

Thus we can apply our inductive assumption (with $\bar{f}$ instead of f), and there exists a $(M, s/2)$-ultra-differentiable symplectic transformation

$$\Psi_{j-1} : \mathbb{T}^n \times D_{1/2} \to \mathbb{T}^n \times D_{\delta_{j-1}}$$

such that

$$(H \circ \Psi^j - \hat{f}) \circ \Psi_{j-1} = L_{j-1} + g_{j-1} + \tilde{\bar{f}} + \bar{f}^+ = L_j + g_j + \tilde{\bar{f}} + \bar{f}^+$$

with $\{\tilde{\bar{f}}, L_i\} = 0$ for $1 \leq i \leq j - 1$ and the estimates

$$|\Psi_{j-1} - \mathrm{Id}|_{s/2} \prec \max_{1 \leq i \leq j-1}\{T_i \nu\}, \quad |\tilde{\bar{f}}|_{s/2} \prec \nu, \quad |\bar{f}^+|_{s/2} \prec \nu \exp\left(-\kappa_2 \left(C^{-1}(Ks)\right)^{-1}\right).$$

Moreover, if I is an integrable Hamiltonian such that $\{\bar{f}, I\} = 0$, then $\{\tilde{\bar{f}}, I\} = 0$. We may then set

$$\Psi_j := \Psi^j \circ \Psi_{j-1}, \quad \tilde{f} := \tilde{\bar{f}}, \quad f^+ := \bar{f}^+ + \hat{f} \circ \Psi_{j-1}.$$

Obviously we have

$$|\tilde{f}|_{s/2} \prec \nu.$$

Using Proposition 16, we get

$$|\Psi_j - \mathrm{Id}|_{s/2} \leq |\Psi^j - \mathrm{Id}|_{s_{j-1}} + |\Psi_{j-1} - \mathrm{Id}|_{s/2} \prec \max_{1 \leq i \leq j}\{T_i \nu\}$$

but also

$$|f^+|_{s/2} \leq |\bar{f}^+|_{s/2} + |\hat{f} \circ \Psi_{j-1}|_{s/2} \leq |\bar{f}^+|_{s/2} + |\hat{f}|_{s_{j-1}} \prec \nu \exp\left(-\kappa_2 \left(C^{-1}(Ks)\right)^{-1}\right).$$

From the first part of the inductive assumption, we know that $\{\tilde{f}, L_i\} = 0$ for $1 \leq i \leq j - 1$, but since $\{\bar{f}, L_j\} = 0$, from the second part of the inductive assumption (applied to $S = L_j$) we obtain $\{\tilde{f}, L_j\} = 0$ and hence $\{\tilde{f}, L_i\} = 0$ for $1 \leq i \leq j$. This proves the first part of the statement, but the second part is obvious, as $\{f, I\} = 0$ implies $\{\bar{f}, I\} = 0$ which implies $\{\tilde{f}, I\} = 0$. □

Now let us come back to a Hamiltonian as in $(*)$. For any $1 \leq j \leq n$, let $I_j \in D_{1/2}$, and as before, we still consider a T_j-periodic vector v_j and a positive parameter ρ_j.

PROPOSITION 43. *Let $1 \leq j \leq n$ and H be as in $(*)$, and given $K \geq 1$, assume that*

(142)
$$\begin{cases} |\nabla h(I_j) - v_j| \leq \rho_j, & \rho_j \cdot< 1, \\ \varepsilon \cdot< \rho_i^2, & T_i \rho_i \cdot< s^2, \quad K T_i \rho_i \cdot< 1, \quad 1 \leq i \leq j, \\ |\omega_j - \omega_{j-1}| \cdot< \rho_{j-1}, & \rho_j \cdot< \rho_{j-1}, \quad 2 \leq j. \end{cases}$$

Then there exists a smooth symplectic transformation

$$\Phi_j : \mathbb{T}^n \times D_{\rho_j/2}(I_j) \to \mathbb{T}^n \times D_{\delta_j \rho_j}(I_j)$$

such that

$$H \circ \Phi_j = h + f' + f^*$$

with $\{f', L_i\} = 0$ for $1 \leq i \leq j$ and with the estimates

(143)
$$\begin{cases} |\Pi_I \Phi_j - \mathrm{Id}_I|_{C^0(\mathbb{T}^n \times D_{\rho_j/2}(I_j))} \cdot< \max_{1 \leq i \leq j} \{T_i\} \varepsilon, \\ |\partial_\theta f^*|_{C^0(\mathbb{T}^n \times D_{\rho_j/2}(I_j))} \cdot< \varepsilon s^{-1} \exp\left(-\kappa_2 \left(C^{-1}(Ks)\right)^{-1}\right). \end{cases}$$

Let us point out that one can obtain better (anisotropic) estimates such as those obtained in Proposition 32 or Proposition 33; yet only the estimates (143) will be useful in the sequel.

PROOF. Observe that the assumption $\rho_j \cdot< 1$ ensures that $D_{\delta_j \rho_j}(I_j)$ is contained in D. Without loss of generality, we may assume that $I_j = 0$ and $h(0) = 0$, and we set $\omega := \nabla h(0)$. Let $\sigma_j(\theta, I) := (\theta, \rho_j I)$: expanding $h \circ \sigma_j$ at order 2 around $I = 0$ we have

$$h \circ \sigma_j(I) = \rho_j \omega \cdot I + \rho_j^2 h_j(I), \quad I \in D_{\delta_j},$$

with $|\nabla h_j|_{s/2} \cdot< |\nabla h|_{s/2} \cdot< s^{-1}$. Now consider $H_j := \rho_j^{-1} H \circ \sigma_j$; it can be written as

$$H_j(\theta, I) = \omega \cdot I + \rho_j h_j(I) + \rho_j^{-1} f \circ \sigma_j(\theta, I) = \omega_j \cdot I + g_j(I) + f_j$$

with

$$g_j(I) := (\omega - v_j) \cdot I + \rho_j h_j(I), \quad f_j(I) : \rho_j^{-1} f \circ \sigma_j(\theta, I)$$

satisfying the estimates

$$|g_j|_{s/2} \leq s^{-1} \rho_j, \quad |g_j|_{s/2} \leq \rho_j^{-1} \varepsilon.$$

To this Hamiltonian we can then apply Proposition 42, with $s/2$ instead of $2^{j-1} s$, $s^{-1} \rho_j$ instead of ρ_j and $\nu = \rho_j^{-1} \varepsilon$, as (142) implies (140): for $s' \cdot= s$, we obtain a (M, s')-ultra differentiable symplectic transformation

$$\Psi_j : \mathbb{T}^n \times D_{1/2} \to \mathbb{T}^n \times B_{\delta_j}$$

such that

$$H_j \circ \Psi_j = L_j + g_j + \tilde{f} + f^+$$

with $\{\tilde{f}, L_i\} = 0$ for $1 \leq i \leq j$ and Ψ_j and f^+ satisfying the estimates (141) with $s^{-1} \rho_j$ instead of ρ_j and $\nu = \rho_j^{-1} \varepsilon$. We can eventually set

$$\Phi_j := \sigma_j \circ \Psi_j \circ \sigma_j^{-1}, \quad f' := \rho_j \tilde{f} \circ \sigma_j^{-1}, \quad f^* := \rho_j f^+ \circ \sigma_j^{-1}.$$

One easily check that $H \circ \Phi_j = h + f' + f^*$, that $\{f', L_i\} = 0$ for $1 \leq i \leq j$ and the wanted estimates (143) on Φ_j and f^* follow easily from the estimates on Ψ_j and f^+ given by (141). $\square$

For $1 \leq j \leq n$, let $\Lambda_j \subseteq \mathbb{R}^n$ be the real subspace orthogonal to $v_1, \ldots, v_j$: since these vectors are linearly independent, Λ_j has dimension $n - j$. Let also Π_{Λ_j} (respectively $\Pi_{\Lambda_j^\perp}$) be the orthogonal projection onto Λ_j (respectively $\Lambda_j^\perp$).

PROPOSITION 44. *Let $H_j := H \circ \Phi_j$ be the Hamiltonian given by Proposition 43, and let $(\theta(t), I(t))$ be a solution of the system associated to H_j with $I(0) \in D_{\rho_j/4}(I_j)$. Then for $1 \leq j \leq n - 1$, we have the following dichotomy:*

(1) *either $I(t) \in D_{\rho_j/2}(I(0))$ for $0 \leq t \leq s \exp\left(\kappa_2 \left(C^{-1}(Ks)\right)^{-1}\right)$,*

(2) *or there exists a time* $0 < t^+ \leq s \exp\left(\kappa_2 \left(C^{-1}(Ks)\right)^{-1}\right)$ *for which* $|I(t^+) - I(0)| = \rho_j/2$ *and*
$$|I(t) - I(0)| \leq \rho_j/2, \quad |\Pi_{\Lambda_j^\perp}(I(t) - I(0))| \lessdot \varepsilon, \quad 0 \leq t \leq t^+.$$

For $j = n$, *we have* $I(t) \in D_{\rho_n/2}(I(0))$ *for* $0 \leq t \leq s \exp\left(-\kappa_2 \left(C^{-1}(Ks)\right)^{-1}\right)$.

PROOF. First assume $1 \leq j \leq n-1$. Let $t^+ > 0$ be the time of first exit of $I(t)$ from $D_{\rho_j/2}(I(0))$. Then either $t^+ > s \exp\left(\kappa_2 \left(C^{-1}(Ks)\right)^{-1}\right)$, in which case (1) clearly holds true, or $t^+ \leq s \exp\left(\kappa_2 \left(C^{-1}(Ks)\right)^{-1}\right)$, in which case we will show that (2) holds true. The facts that $|I(t^+) - I(0)| = \rho_j/2$ and $|I(t) - I(0)| \leq \rho_j/2$ for $0 \leq t \leq t^+$ are obvious. Then since $H \circ \Phi_j = h + f' + f^*$ with $\{f', L_i\} = 0$ for $1 \leq i \leq j$, we have $\partial_\theta f' \in \Lambda_j$, and thus for $0 \leq t \leq t^+$,

$$|\Pi_{\Lambda_j^\perp}(I(t) - I(0))| \leq |t| |\partial_\theta f^*|_{C^0(\mathbb{T}^n \times B_{\rho_j}(I_j))} \lessdot t^+ \varepsilon s^{-1} \exp\left(-\kappa_2 \left(C^{-1}(Ks)\right)^{-1}\right) \lessdot \mu.$$

For $j = n$, since $\{f', L_j\} = 0$ for $1 \leq j \leq n$, the Hamiltonian f' is integrable and the last estimate shows that $t^+ \geq s \exp\left(\kappa_2 \left(C^{-1}(Ks)\right)^{-1}\right)$; then the conclusion follows from $\mu \lessdot \rho_n$. $\square$

Proposition 43 and Proposition 44, which consist respectively in the construction of normal forms and their dynamical consequences, constitute the analytical part of the proof. The geometrical part will be based on two propositions. The first one is Proposition 35, which we already used in § 5 and is a mere consequence of Dirichlet's box principle. The second one is a consequence of the steepness property of h; the statement below is a special case of the lemma on "almost plane curves" of Nekhoroshev, stated in [**Nek77**] and proved in [**Nek79**] (our case below corresponds to "plane curves").

PROPOSITION 45. *Let* $h : G \to \mathbb{R}$ *be a smooth function which is* (l, L, δ, p)-*steep, and such that* $\sup_{I \in G} |\nabla^2 h(I)| \lessdot 1$. *Let* $\gamma : [0, t^*] \to \mathbb{R}^n$ *be a continuous curve,* λ *a proper affine subspace of* $\mathbb{R}^n$ *and* $\varrho > 0$. *Assume that*

 (i) *for all* $t \in [0, t^*]$, $\gamma(t) \in \lambda$;
 (ii) *for all* $t \in [0, t^*]$, $|\gamma(0) - \gamma(t)| \leq \varrho$ *and* $|\gamma(0) - \gamma(t^*)| = \varrho$;
 (iii) *the ball* $\{I \in \mathbb{R}^n \mid |I - \gamma(0)| \leq \varrho\}$ *is contained in* G;
 (iv) $\varrho \lessdot 1$,

then there exists a time $\tilde{t} \in [0, t^*]$ *such that*

$$\|\Pi_\Lambda \nabla h(\gamma(\tilde{t}))\| \gtrdot \varrho^p,$$

where Λ *is the vector space associated to* λ.

We can now conclude the proof of Theorem M.

PROOF OF THEOREM M. For $1 \leq j \leq n$, let us set

$$a_j := (np)^{n-j}, \quad Q_j := \varepsilon^{-\frac{1}{2na_j}}$$

and

$$K := Q_1 := \varepsilon^{-\frac{1}{2na_1}}.$$

Given any solution $(\theta(t), I(t))$ of the Hamiltonian system associated to H with $I(0) \in D_{1/2}$, we claim that

$$|I(t) - I(0)| \lessdot Q_1^{-1}, \quad |t| \leq s \exp\left(\kappa_2 \left(C^{-1}(Ks)\right)^{-1}\right)$$

provided that

(144) $$\varepsilon \lessdot 1, \quad Q_1^{-1} \lessdot s^2.$$

which holds true if ε is sufficiently small.

In view of the definitions of Q_1, K and the fact that $a_1 = a$, this claim clearly implies the statement we want to prove for positive times; for negative times the argument is entirely similar.

So it remains to prove the claim, and we will do this by an algorithm that stops after at most n steps. Let us write

$$\tau_K := s \exp\left(\kappa_2 \left(C^{-1}(Ks)\right)^{-1}\right).$$

For the first step, we set $I_1 := I(0)$ and we apply Proposition 35 to $\omega := \nabla h(I_1)$ with $Q = Q_1$: we find a T_1-periodic vector $v_1 \in \mathbb{R}^n \setminus \{0\}$ such that

$$|\nabla h(I_1) - v_1| \lessdot (T_1 Q_1)^{-1}, \quad 1 \lessdot T_1 \lessdot Q_1^{n-1}.$$

Let us set $\rho_1 = \cdot (T_1 Q_1)^{-1}$. The conditions (144) imply (142) for $j = 1$ and hence Proposition 43 can be applied: let $H_1 := H \circ \Phi_1$ be the new Hamiltonian, and let us denote by $(\theta^1(t), I^1(t))$ its solutions. Consider the solution for which $\Phi_1((\theta^1(0), I^1(0))) = (\theta(0), I(0))$: in view of the estimate on Φ_1 given in Proposition 43, $I^1(0) \in D_{\rho_1/4}(I_1)$ and Proposition 44 gives a dichotomy:

(1) either $I^1(t) \in D_{\rho_1/2}(I^1(0))$ for $0 \le t \le \tau_K$,
(2) or there exists a time $0 < t_1^+ \le \tau_K$ for which $|I^1(t_1^+) - I^1(0)| = \rho_1/2$ and

$$|I^1(t) - I^1(0)| \le \rho_1/2, \quad |\Pi_{\Lambda_1^\perp}(I^1(t) - I^1(0))| \lessdot \varepsilon, \quad 0 \le t \le t_1^+.$$

If the first alternative holds true, the claim is proved since this implies

$$|I(t) - I(0)| \le 2\rho_1 \lessdot Q_1^{-1}, \quad 0 \le t \le \tau_K.$$

If the second alternative holds true, we will show how the algorithm moves to the second step. Consider the curve

$$\gamma_1(t) := I^1(0) + \Pi_{\Lambda_1}\left(I^1(t) - I^1(0)\right).$$

Since $\varepsilon \cdot \lessdot \rho_1$, we can ensure that

$$|\gamma_1(t_1^+) - \gamma_1(0)| = |\Pi_{\Lambda_1}\left(I^1(t_1^+) - I^1(0)\right)| \ge |I^1(t_1^+) - I^1(0)| - |\Pi_{\Lambda_1^\perp}(I^1(t_1^+) - I^1(0))| \ge \rho_1/4$$

and therefore we can find a time $0 < t_1^* < t_1^+$ for which

$$\begin{cases} |\gamma_1(t) - \gamma_1(0)| \le \rho_1/4, & 0 \le t \le t_1^*, \\ |\gamma_1(t_1^*) - \gamma_1(0)| = \rho_1/4. \end{cases}$$

This curve γ_1, on the interval $[0, t_1^*]$, satisfies the assumption of Proposition 45 with $\lambda := I^1(0) + \Lambda_1$, $\varrho = \rho_1/4$ and therefore we can find a time $\tilde{t}_1 \in [0, t_1^*]$ such that

$$|\Pi_{\Lambda_1}\nabla h(\gamma_1(\tilde{t}_1))| \cdot > \rho_1^p.$$

Moreover,

$$|\gamma_1(\tilde{t}_1) - I^1(\tilde{t}_1)| = |\Pi_{\Lambda_1^\perp}\left(I^1(\tilde{t}_1) - I^1(0)\right)| \lessdot \varepsilon$$

and since $\varepsilon \cdot \lessdot \rho_1^p$ as one easily check, we find that

$$|\Pi_{\Lambda_1}\nabla h(I^1(\tilde{t}_1))| \cdot > \rho_1^p.$$

Then, using the estimate on Φ_1 given in Proposition 43 and the fact that $T_1\varepsilon \cdot \lessdot \rho_1^p$, one also has

$$|I(\tilde{t}_1) - I^1(\tilde{t}_1)| \cdot \lessdot \rho_1^p$$

and thus

$$|\Pi_{\Lambda_1}\nabla h(I(\tilde{t}_1))| \cdot > \rho_1^p.$$

Now we set $I_2 := I(\tilde{t}_1)$, and we apply again Proposition 35 to $\omega := \nabla h(I_2)$ with $Q = Q_2$: we find a T_2-periodic vector $v_2 \in \mathbb{R}^n \setminus \{0\}$ such that

$$|\nabla h(I_2) - v_2| \lessdot (T_2 Q_2)^{-1}, \quad 1 \lessdot T_2 \lessdot Q_2^{n-1}.$$

Let us set $\rho_2 = \cdot (T_2 Q_2)^{-1}$. To check that v_2 is linearly independent from v_1, observe that

$$\rho_2 \lessdot Q_2^{-1} \cdot \lessdot Q_1^{-np} \cdot \lessdot \rho_1^p$$

and therefore, by a proper choice of implicit constants, one can ensure that

$$|\Pi_{\Lambda_1}v_2| \ge |\Pi_{\Lambda_1}\nabla h(I_2))| - |\Pi_{\Lambda_1}\left(\nabla h(I_2) - v_2\right)| \ge |\Pi_{\Lambda_1}\nabla h(I_2))| - |\nabla h(I_2) - v_2| > 0.$$

Clearly, we have $\rho_2 \cdot \lessdot \rho_1$ and it is straightforward to check that $|v_1 - v_2| \lessdot \rho_1$. Then conditions (144) imply (142) for $j = 2$, and once again, we can apply Proposition 43 and Proposition 44 leading

to a new dichotomy: if the first alternative holds true, the claim is proved while if the second alternative holds true, we move to the next step. If the algorithm lasts $n - 1$ steps, then only possibility (1) in Proposition 44 can hold and this concludes the proof of the claim. $\square$

On the moderate growth condition

Recall from § 1 that the sequence M has *moderate growth* if

(MG) $$\sup_{l,j\in\mathbb{N}}\left(\frac{M_{l+j}}{M_l M_j}\right)^{\frac{1}{l+j}} < +\infty$$

Here we compare this condition to (H2), which requires the sequence $\left(\frac{\ln\mu_l}{l}\right)_{l\geq 1}$ to converge to zero as l goes to infinity; in particular, (H2) is satisfied if $\left(\frac{\ln\mu_l}{\ln l}\right)_{l\geq 2}$ is bounded.

LEMMA 46. *If conditions* (H1) *and* (MG) *hold, positive sequences* $\left(\frac{\ln M_l}{l\ln l}\right)_{l\geq 2}$ *and* $\left(\frac{\ln\mu_l}{\ln l}\right)_{l\geq 2}$ *are bounded away from* 0 *and* $+\infty$ *and, in particular,* (H2) *is satisfied.*

PROOF. As noticed in § 2, (H1) implies

$$M_l \geq l!, \quad \mu_l \geq l+1 \quad (\forall l\in\mathbb{N}),$$

which shows that both sequences of the statement are lower bounded.

Now, assume that (M_l) has moderate growth. In restriction to $j=l$, (MG) reduces to

$$M_{2l} \leq A^l M_l^2$$

for some $A\in[1,+\infty[$ independent of $l\in\mathbb{N}$. Using that $M_1 = 1$, by induction we see that

(145) $$M_{2^k} \leq A^{k2^{k-1}} \quad (k\in\mathbb{N}).$$

As already noticed, (H1) implies that M itself is log-convex, hence, if $2^k \leq l < 2^{k+1}$,

$$M_l \leq M_{2^k}^{\frac{2^{k+1}-l}{2^k}} M_{2^{k+1}}^{\frac{l-2^k}{2^k}} \leq \left(M_{2^k}M_{2^{k+1}}\right)^{\frac{l}{2^k}}.$$

Using (145), if $k\geq 1$,

$$M_l \leq A^{lk/2+l(k+1)} \leq A^{5lk/2},$$

whence

$$M_l \leq \left(A^{\frac{5}{2\ln 2}}\right)^{l\ln l} \quad (\forall l\geq 2),$$

and the sequence $\frac{\ln M_l}{l\ln l}$ is bounded.

Let us now see why $\rho_l := \frac{\ln\mu_l}{\ln l}$ is upper bounded. For any given $l\in\mathbb{N}$, let $k = k_l\in\mathbb{N}$ be such that $2^{k-1}\leq l < 2^k$. By the definition (4) of μ,

$$M_{2^{k+1}} = M_{2^k}\mu_{2^k}\cdots\mu_{2^{k+1}-1}.$$

Because of (H1), the sequence (M_{2^k}) is larger than 1 and the sequence μ is increasing, so

$$M_{2^{k+1}} \geq (\mu_{2^k})^{2^k} \geq (\mu_l)^{2^k} = e^{2^k\rho_l\ln l}.$$

Using (145), we get

$$\rho_l\ln l \leq (k+1)\ln A,$$

and as

$$(k+1) \leq \frac{\ln l}{\ln 2} + 2$$

we have

$$\rho_l \leq \ln A\left(\frac{1}{\ln 2} + \frac{2}{\ln l}\right) \quad (\forall l\geq 1),$$

which shows that $(\rho_l)_{l\geq 1}$ must be bounded. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

Bibliography

[Arn64]   V.I. Arnold, *Instability of dynamical systems with several degrees of freedom*, Sov. Math. Doklady **5** (1964), 581–585.

[Bes96]   U. Bessi, *An approach to Arnold's diffusion through the calculus of variations*, Nonlinear Anal., Theory Methods Appl. **26** (1996), no. 6, 1115–1135.

[Bes97]   _____, *Arnold's example with three rotators*, Nonlinearity **10** (1997), no. 3, 763–781.

[Bes00]   _____, *An analytic counterexample to the KAM theorem*, Ergodic Theory Dynam. Systems **20** (2000), no. 2, 317–333.

[BF13]   A. Bounemoura and S. Fischler, *A diophantine duality applied to the KAM and Nekhoroshev theorems*, Math. Z. **275** (2013), no. 3, 1135–1167.

[BF14]   _____, *The classical KAM theorem for Hamiltonian systems via rational approximations*, Regul. Chaotic Dyn. **19** (2014), no. 2, 251–265.

[BF17]   A. Bounemoura and J. Féjoz, *KAM, α-Gevrey regularity and the α-Bruno-Rüssmann condition*, preprint, 2017.

[BFN17]   A. Bounemoura, B. Fayad, and L. Niederman, *Superexponential stability of quasi-periodic motion in Hamiltonian systems*, Comm. Math. Phys. **350** (2017), no. 1, 361–386.

[BM11]   A. Bounemoura and J.-P. Marco, *Improved exponential stability for near-integrable Hamiltonian systems*, Nonlinearity **24** (2011), no. 1, 97–112.

[BN12]   A. Bounemoura and L. Niederman, *Generic Nekhoroshev theory without small divisors*, Ann. Inst. Fourier **62** (2012), no. 1, 277–324.

[Bou11]   A. Bounemoura, *Effective stability for Gevrey and finitely differentiable prevalent Hamiltonians*, Communication in Mathematical Physics **307** (2011), no. 1, 157–183.

[Bou12]   _____, *Optimal stability and instability for near-linear hamiltonians*, Annales Henri Poincaré **13** (2012), no. 4, 857–868.

[Bou13a] _____, *Normal forms, stability and splitting of invariant manifolds I. Gevrey Hamiltonians*, Regul. Chaotic Dyn. **18** (2013), no. 3, 237–260.

[Bou13b] _____, *Normal forms, stability and splitting of invariant manifolds I. Gevrey Hamiltonians*, Regul. Chaotic Dyn. **18** (2013), no. 3, 237–260.

[Bru71]   A.D. Bruno, *Analytical form of differential equations I*, Trans. Moscow Math. Soc. **25** (1971), 131–288.

[Bru72]   _____, *Analytical form of differential equations II*, Trans. Moscow Math. Soc. **26** (1972), 199–239.

[Car03]   T. Carletti, *The Lagrange inversion formula on non-Archimedean fields. Non-analytical form of differential and finite difference equations*, Discrete Contin. Dyn. Syst. **9** (2003), no. 4, 835–858.

[CC94]   J. Chaumat and A.-M. Chollet, *Surjectivité de l'application restriction à un compact dans des classes de fonctions ultradifférentiables*, Math. Ann. **298** (1994), no. 1, 7–40.

[CM00]   T. Carletti and S. Marmi, *Linearization of analytic and non-analytic germs of diffeomorphisms of* (**C**, 0), Bull. Soc. Math. France **128** (2000), no. 1, 69–85.

[CW13]   C.-Q. Cheng and L. Wang, *Destruction of Lagrangian torus for positive definite Hamiltonian systems*, Geom. Funct. Anal. **23** (2013), no. 3, 848–866.

[Fas90]   F. Fasso, *Lie series method for vector fields and Hamiltonian perturbation theory*, Z. Angew. Math. Phys. **41** (1990), no. 6, 843–864 (English).

[Fay02]   B. R. Fayad, *Weak mixing for reparameterized linear flows on the torus*, Ergodic Theory Dynam. Systems **22** (2002), no. 1, 187–201.

[GCB16]   M. Guzzo, L Chierchia, and G. Benettin, *The steep Nekhoroshev's theorem*, Comm. Math. Phys. **342** (2016), 569–601.

[Gev18]   M. Gevrey, *Sur la nature analytique des solutions des équations aux dérivées partielles. Premier mémoire.*, Ann. Sci. Ec. Norm. Sup. **35** (1918), 129–190.

[Kar16]   N. Karaliolios, *Local rigidity of Diophantine translations in higher dimensional tori*, preprint, 2016.

[Kol54]   A.N. Kolmogorov, *On the preservation of conditionally periodic motions for a small change in Hamilton's function*, Dokl. Akad. Nauk. SSSR **98** (1954), 527–530.

[Kom79]   H. Komatsu, *The implicit function theorem for ultradifferentiable mappings*, Proc. Japan Acad. Ser. A Math. Sci. **55** (1979), no. 3, 69–72.

[Kom80]   _____, *Ultradifferentiability of solutions of ordinary differential equations*, Proc. Japan Acad. Ser. A Math. Sci. **56** (1980), no. 4, 137–142.

[Lax53]   P. D. Lax, *Nonlinear hyperbolic equations*, Comm. Pure Appl. Math. **6** (1953), 231–258.

[LN92]   P. Lochak and A.I. Neishtadt, *Estimates of stability time for nearly integrable systems with a quasiconvex Hamiltonian*, Chaos **2** (1992), no. 4, 495–499.

[LNN94]   P. Lochak, A.I. Neistadt, and L. Niederman, *Stability of nearly integrable convex Hamiltonian systems over exponentially long times*, Kuksin, S. (ed.) et al., Seminar on dynamical systems. Basel: Birkhäuser. Prog. Nonlinear Differ. Equ. Appl. 12, 15-34 (1994)., 1994.

[Loc92]   P. Lochak, *Canonical perturbation theory via simultaneous approximation*, Russ. Math. Surv. **47** (1992), no. 6, 57–133.

[MS02]   J.-P. Marco and D. Sauzin, *Stability and instability for Gevrey quasi-convex near-integrable Hamiltonian systems*, Publ. Math. Inst. Hautes Études Sci. **96** (2002), 199–275.

[Nek77]   N.N. Nekhoroshev, *An exponential estimate of the time of stability of nearly integrable Hamiltonian systems*, Russian Math. Surveys **32** (1977), no. 6, 1–65.

[Nek79]   _____, *An exponential estimate of the time of stability of nearly integrable Hamiltonian systems II*, Trudy Sem. Petrovs **5** (1979), 5–50.

[Pop04]   G. Popov, *KAM theorem for Gevrey Hamiltonians*, Erg. Th. Dyn. Sys. **24** (2004), no. 5, 1753–1786.

[Pös93]   J. Pöschel, *Nekhoroshev estimates for quasi-convex Hamiltonian systems*, Math. Z. **213** (1993), 187–216.

[Pös01]   _____, *A lecture on the classical KAM theory*, Katok, Anatole (ed.) et al., Smooth ergodic theory and its applications (Seattle, WA, 1999). Providence, RI: Amer. Math. Soc. (AMS). Proc. Symp. Pure Math. 69, 707-732, 2001.

[RS96]   J.-P. Ramis and R. Schäfke, *Gevrey separation of fast and slow variables*, Nonlinearity **9** (1996), no. 2, 353–384.

[Rüs75]   H. Rüssmann, *On optimal estimates for the solutions of linear partial differential equations of first order with constant coefficients on the torus*, Dynamical systems, theory and applications (Rencontres, Battelle Res. Inst., Seattle, Wash., 1974), Springer, Berlin, 1975, pp. 598–624.Lecture Notes in Phys., Vol. 38.

[Rüs01]   _____, *Invariant tori in non-degenerate nearly integrable Hamiltonian systems*, Regul. Chaotic Dyn. **6** (2001), no. 2, 119–204.

[SCK03]   C.E. Shin, S.-Y. Chung, and D. Kim, *Gevrey and analytic convergence of Picard's successive approximations*, Integral Transforms Spec. Funct. **14** (2003), no. 1, 19–30.

[Thi03]   V. Thilliez, *Division by flat ultradifferentiable functions and sectorial extensions*, Results Math. **44** (2003), no. 1-2, 169–188.

[Thi08]   _____, *On quasianalytic local rings*, Expo. Math. **26** (2008), no. 1, 1–23.

[TZ10]   D. Treschev and O. Zubelevich, *Introduction to the perturbation theory of Hamiltonian systems*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2010.

[Vil08]   J. Villanueva, *Kolmogorov theorem revisited*, J. Differential Equations **244** (2008), no. 9, 2251–2276.

[Yoc88]   J.-C. Yoccoz, *Linéarisation des germes de difféomorphismes holomorphes de* $(\mathbf{C}, 0)$, C. R. Acad. Sci. Paris Sér. I Math. **306** (1988), no. 1, 55–58. MR 929279

[Yoc95]   _____, *Small divisors in dimension one (Petits diviseurs en dimension 1)*, Astérisque. 231. Paris: Société Math. de France, 242 p., 1995.

[Zha11]   Ke Zhang, *Speed of Arnold diffusion for analytic Hamiltonian systems*, Inventiones mathematicae **186** (2011), no. 2, 255–290.

[ZZ17]   J. Zhang and K. Zhang, *Improved stability for analytic quasi-convex nearly integrable systems and optimal speed of arnold diffusion*, preprint, 2017.