A parametrised version of Moser's modifying terms theorem

Wagener, F.

Publication date
2009

Document Version
Submitted manuscript

Manuscript submitted to
AIMS' Journals
Volume **X**, Number **0X**, XX **200X**

Website: http://AIMsciences.org
pp. **X–XX**

A PARAMETRISED VERSION OF MOSER'S MODIFYING TERMS THEOREM

Abstract. A sharpened version of Moser's 'modifying terms' KAM theorem is derived, and it is shown how this theorem can be used to investigate the persistence of invariant tori in general situations, including those where some of the Floquet exponents of the invariant torus may vanish. The result is 'structural' and works for dissipative, Hamiltonian, reversible and symmetric vector fields. These results are derived for the contexts of real analytic, Gevrey regular, ultradifferentiable and finitely differentiable perturbed vector fields. In the first two cases, the conjugacy constructed in the theorem is shown to be Gevrey smooth in the sense of Whitney on the set of parameters satisfy a "Diophantine" non-resonance condition.

Florian Wagener

CeNDEF, Dept. of Quantitative Economics, University of Amsterdam
Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

1. Introduction.

1.1. **Object.** Moser's modifying terms theorem [23] is in essence an averaging result. On the phase space $\mathscr{M} = \mathbb{T}^m \times \mathbb{R}^n$, it considers small deformations $\tilde{X}$ of an integrable vector field

$$X = \omega \frac{\partial}{\partial x} + Ay\frac{\partial}{\partial y}, \qquad x \in \mathbb{T}^m,\, y \in \mathbb{R}^n, \tag{1}$$

where $\omega \in \mathbb{R}^m$ and $A \in \mathsf{gl}(n,\mathbb{R})$ are constant and are assumed to satisfy so-called *Diophantine non-resonance conditions.* The theorem says that if the deformation is sufficiently small in some function norm, say

$$0 < \varepsilon = \|\tilde{X} - X\| \ll 1$$

then there is a constant vector field

$$\Delta = \delta\frac{\partial}{\partial x} + (\mu + By)\frac{\partial}{\partial y},$$

with $\delta \in \mathbb{R}^m$, $\mu \in \mathbb{R}^m$ and $B \in \mathsf{gl}(n,\mathbb{R})$, such that the following holds. If X_0 denotes the *modified* vector field

$$X_0 = \tilde{X} - \Delta,$$

2000 *Mathematics Subject Classification.* Primary: 37C55; Secondary: 37G99.

Key words and phrases. Quasi-periodic tori, modifying terms, normal-internal resonances.

The author thanks Henk Broer, Francesco Fassò, Heinz Hanßmann, Jun Hoo, George Huitema and Mikhail Sevryuk for fruitful discussions and helpful remarks. This work has been supported by the Netherlands Organisation for Science (NWO) under the CeNDEF PIONIER grant and under a MaGW-VIDI grant.

then there is a conjugacy Φ, ε-close to the identity, for which

$$\Phi_* X_0 = \Big(\omega + \mathsf{O}(|y|)\Big)\frac{\partial}{\partial x} + \Big(Ay + \mathsf{O}(|y|^2)\Big)\frac{\partial}{\partial y}. \tag{2}$$

In particular, the torus $\mathscr{T} = \mathbb{T}^m \times \{0\}$ is invariant under $\Phi_* X_0$, and consequently the torus $\Phi^{-1}(\mathscr{T})$ is invariant under X_0. The natural interpretation of the vector field Δ is that it represents that part of the perturbation which cannot be removed by successive averaging.

The object of the present article is to derive a modifying terms theorem for parametrised families of vector fields, incorporating results on smoothness [28] and Gevrey-regularity [25, 26] of parameter dependence that have been added to KAM-theory since Moser's article appeared. An extension to general Carleman (or ultra-differentiable) classes is given as well. A second motivation is to make the result a convenient tool for quasi-periodic bifurcation theory. In particular, the condition imposed by Moser that ad_A should be semi-simple is removed, so that all situations can be treated for which the unperturbed invariant tori have several Floquet exponents equal to zero. Recall that if a vector field is of the form of the right hand side of equation (2), then the eigenvalues of A are called the *Floquet exponents* of the invariant torus $\mathscr{T}$. As an application, we sketch the analysis of persistence of tori in the quasi-periodic Bogdanov-Takens bifurcation.

The main result of the present article is to show the existence of a modifying terms vector field Δ with the above properties, for small parametrised deformations $\tilde{X}$ of integrable vector fields X. Here the vector fields X and $\tilde{X}$ can be restricted to an admissible structure in the sense of [9], like Hamiltonian, volume preserving, equivariant etc. The deformations are either real analytic, Gevrey-regular, ultradifferentiable or finitely (but sufficiently often) differentiable, and for each category we find regularity properties of the conjugacy Φ and the vector field Δ. In this way, the results contribute to a resolution of problem 10 of Sevryuk's list [33].

1.2. **Related work.** Invariant tori with one or more vanishing Floquet exponents occur in the integrable versions of many bifurcation scenarios. In the context of a degenerate Hopf bifurcation Chenciner [13] has investigated the saddle-node bifurcation of invariant quasi-periodic circles. His results have been extended by Broer, Huitema, Takens and Braaksma [9], and, in the context of Hamiltonian vector fields, by Hanßmann [18]. The scope of these studies is restricted to the case of a one-dimensional normal space, in the general context, or a two-dimensional normal space, in the Hamiltonian context. More recently, higher order degeneracies have been studied as well [6, 19, 37].

For one-dimensional normal spaces, the Rüssmann-Herman translated torus theorem is available, which is the discrete-time analogon of the modifying terms theorem. Recently, there modifying terms theorem has been applied in several settings [15, 16].

Higher dimensional normal spaces have been treated extensively by other methods in the case of non-vanishing Floquet exponents; we refer the reader to [8, 9] and the references there. The results reached in those investigations were restricted to the case that all Floquet exponents are distinct; recently, this restriction has been removed by the work of Hoo [20], which extended previous work of de Jong [17] and Ciocci [14].

Pöschel [28] demonstrated that the conjugacies of KAM theory depend differentiable in the sense of Whitney on the parameters, even in the case that the original deformation is only finitely often differentiable. This has been strengthened by Popov [25, 26] to Gevrey-regularity in the sense of Whitney if the deformation $\tilde{X}$ itself is real analytic (cf. also [12]). Popov [27] extended these results to Hamiltonians that are Gevrey regular. Much subsequent work has been done, taking especially into account the Rüssmann condition [38, 40, 41, 43, 42]. A simple derivation of this kind of results has been given in [36]; the same method is used to obtain the results on ultradifferentiable deformations in the present article.

1.3. **Structure of the article.** The next section, after introducing notation, states the central 'KAM-averaging' theorem (theorem 2.3). In section 3 it is shown how the results of [23] and [9] are corollaries of the theorem; moreover, a quasi-periodic analogue of Arnol'd's succinct 'persistence of bifurcation' result [1] is derived. The proof of the central theorem occupies section 4.

2. **Modifying terms.** This section introduces notations used throughout the article, and states the modifying terms theorem.

2.1. **Notations and definitions.** Let $\langle y_1, y_2\rangle$ denote the standard Euclidean or Hermitian inner product of two vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$, and let $|y|$ denote the norm $|y| = \max_i |y_i|$. Let $\mathbb{T}^m$ be the standard m–torus $\mathbb{R}^m/2\pi\mathbb{Z}^m$.

2.1.1. *Vector fields and invariant tori.* In the following, a *family* of objects is always taken in the sense as a parametrised family, where the parameter takes values in some subset of a finite dimensional vector space.

Let $\mathscr{M}$ be a manifold. We consider small deformations of families of vector fields X on $\mathscr{M}$ that leave a family of embedded tori $\mathscr{T}$ invariant. Let $\mathsf{T}\mathscr{M}$, $\mathsf{T}\mathscr{T}$ and and $\mathsf{T}_{\mathscr{T}}\mathscr{M}$ denote respectively the tangent bundle to $\mathscr{M}$, the tangent bundle to $\mathscr{T}$ and the restriction of $\mathsf{T}\mathscr{M}$ to $\mathscr{T}$. The quotient $\mathsf{T}_{\mathscr{T}}\mathscr{M}/\mathsf{T}\mathscr{T}$ is a smooth vector bundle over $\mathscr{T}$, the *normal bundle* $\mathsf{N}\mathscr{T}$ of $\mathscr{T}$. By the tubular neighbourhood theorem, $\mathsf{N}\mathscr{T}$ is diffeomorphic to an open neighbourhood U of $\mathscr{T}$. Assuming the normal bundel to be trivial, the diffeomorphism transfers vector fields on U to vector fields on $\mathsf{N}\mathscr{T} \cong \mathbb{T}^m \times \mathbb{R}^n$; note that then $\mathscr{T} \cong \mathbb{T}^m \times \{0\}$.

Accordingly, in the following families of vector fields $X(p)$ on the phase space $\mathscr{M} = \mathbb{T}^m \times \mathbb{R}^n$ will be considered, where the parameter p takes values in a space $\mathscr{P}$ which is an open and bounded neighbourhood of the origin of $\mathbb{R}^q$. Note that $\mathscr{M}$ can still be identified with the normal bundle $\mathsf{N}\mathscr{T}$ of the torus $\mathscr{T}$.

A regularly parametrised family of vector fields $p \mapsto X(p)$ is usually not distinguished from the equivalent *vertical* vector field X on $\mathscr{M} \times \mathscr{P}$. Recall that a vector field is called vertical if the canonical projection of X to the tangent bundle $\mathsf{T}\mathscr{P}$ of the parameter space $\mathscr{P}$ vanishes everywhere. A vertical vector field on $\mathscr{M} \times \mathscr{P}$ is typically written as

$$X = f(x,y,p)\frac{\partial}{\partial x} + g(x,y,p)\frac{\partial}{\partial y}, \tag{3}$$

where $x \in \mathbb{T}^m$, $y \in \mathbb{R}^n$ and $p \in \mathscr{P}$. The set of all differentiable vertical vector fields on $\mathscr{M} \times \mathscr{P}$ is denoted by $\mathcal{X}$.

2.1.2. *Normal linear vector fields.* If $X \in \mathcal{X}$ is a vector field of the form (3), the *normal linear part* $\mathsf{N}X$ *of* X is defined as

$$\mathsf{N}X = f(x,0,p)\frac{\partial}{\partial x} + \left(g(x,0,p) + \frac{\partial g}{\partial y}(x,0,p)y\right)\frac{\partial}{\partial y}. \tag{4}$$

Note that the flow of $\mathsf{N}X$ maps fibers of the normal bundle $\mathsf{N}\mathscr{T}$ affinely to fibers; "normal affine vector field" would perhaps be a more appropriate name, but we stick to the convention introduced in [9]. Generally, a vector field L will be called *normally linear* if it is equal to its normal linear part.

If $X \in \mathcal{X}$ is such that the term $g(x, 0, p)$ in (4) vanishes identically, then X is tangent to the torus $\mathscr{T}$, and $\mathscr{T}$ is invariant under the flow of X. Introduce for $\varepsilon > 0$ the scaling diffeomorphism $D_\varepsilon(x, y, p) = (x, \varepsilon^{-1}y, p)$. If X is tangent to $\mathscr{T}$, then $\lim_{\varepsilon\downarrow 0}(D_\varepsilon)_* X = \mathsf{N}X$, and consequently

$$(D_\varepsilon)_* X = \mathsf{N}X + \mathsf{O}(\varepsilon).$$

Hence, without loss of generality, it can be assumed that the unperturbed vertical vector field is normally linear.

2.1.3. *Integrability.* A vertical vector field $X \in \mathcal{X}$ is called *integrable*, if it is equivariant with respect to the action Θ of the group $\mathbb{T}^m$ on $\mathscr{M} \times \mathscr{P}$ that is given as

$$\Theta_\beta(x, y, p) = (x + \beta, y, p)$$

for $\beta \in \mathbb{T}^m$. Equivariance means that

$$(\Theta_\beta)_* X = X$$

for all β. Consequently, if X is integrable, it can be written in the form

$$X = f(y, p)\frac{\partial}{\partial x} + g(y, p)\frac{\partial}{\partial y}. \tag{5}$$

Define the $\mathbb{T}^m$-average $[f]$ of a function f defined on $\mathscr{M} \times \mathscr{P}$ as

$$[f](y, p) = \int_{\mathbb{T}^m} f(x, y, p)\,\mathrm{d}x;$$

here $\mathrm{d}x$ denotes the Haar measure on $\mathbb{T}^m$.

If $X = f\frac{\partial}{\partial x} + g\frac{\partial}{\partial y} \in \mathcal{X}$ is any vector field, the integrable part $[X]$ of X is given as

$$[X] = [f](y, p)\frac{\partial}{\partial x} + [g](y, p)\frac{\partial}{\partial y}.$$

Note that with this definition, a vector field X is integrable if and only if $X = [X]$. A vector field X which is such that $[X] = 0$ is said to be *mean*-0. Any vector field can be decomposed in an integrable part and a mean-0 part:

$$X = [X] + (X - [X]).$$

2.1.4. *Frequencies.* An integrable vector field X of the form (5) can be written uniquely as $X = L + Q$ with $L = \mathsf{N}X$ and $Q = X - \mathsf{N}X$. The normal linear part L of X is then of the form

$$L = \omega(p)\frac{\partial}{\partial x} + (\mu(p) + A(p)y)\,\frac{\partial}{\partial y}. \tag{6}$$

Note that if $\mu(p_0) = 0$, then the vector field $X(p_0)$ is tangent to $\mathscr{T}$, which is consequently invariant.

The maps $\omega : \mathscr{P} \to \mathbb{R}^m$ and $\Omega : \mathscr{P} \to \mathbb{R}^m \times \mathsf{gl}(n, \mathbb{C})$, the latter given by

$$\Omega = (\omega, A),$$

are called the *internal frequency map* and the *(full) frequency map* of X, respectively. For a given frequency map Ω, let

$$L_\Omega = \omega(p)\frac{\partial}{\partial x} + A(p)y\frac{\partial}{\partial y}.$$

2.1.5. *Structures.* In order to describe families vector fields that admit certain symmetries, "admissible structures" are introduced, following [9, 23].

For every $d > 0$ and every vertical vector field X, define the Fourier truncation $T_d X$ of X as

$$T_d X = \sum_{|k|\le d} X_k(y,p)\,\mathrm{e}^{\mathrm{i}\langle k,x\rangle}, \quad \text{where} \quad X_k(y,p) = \int_{\mathbb{T}^m} X(x,y,p)\,\mathrm{e}^{-\mathrm{i}\langle k,x\rangle}\,\mathrm{d}x.$$

An *admissible structure* is a pair $(\mathfrak{g},\mathfrak{h})$, where $\mathfrak{g}$ is the Lie algebra of a finite dimensional Lie group $\mathfrak{G} \subset \mathsf{GL}(n,\mathbb{R})$, and where $\mathfrak{h} \subset \mathcal{X}$ is an infinite dimensional Lie algebra of vector fields on $\mathscr{M}$, such that $\mathfrak{g}$ and $\mathfrak{h}$ satisfy the following properties. For every $X \in \mathfrak{h}$, the normal linear vector field $\mathsf{N}X$ as well as the truncation $T_d X$ is in $\mathfrak{h}$, for every $d > 0$. Moreover, the frequency map $\Omega = (\omega, A)$ of an integrable vector field in $\mathfrak{h}$ takes values in $\mathbb{R}^m \times \mathfrak{g}$.

Let $\mathscr{U}$ be an open and bounded subset of $\mathscr{M}$, and let $\Phi : \mathscr{U} \to \mathscr{M}$ be an embedding. If for any $X \in \mathfrak{h}$ the vector field $\Phi_* X$ is the restriction of a vector field $Y \in \mathfrak{h}$ to $\Phi(U)$, then Φ is called a *structure-preserving* conjugacy associated to $\mathfrak{h}$.

2.1.6. *Versal unfoldings.* A frequency map

$$\bar\Omega = (\bar\omega, \bar A) : \Sigma \to \mathbb{R}^m \times \mathsf{gl}(n,\mathbb{C})$$

is a smooth *versal unfolding* of Ω_0, if for every smooth deformation $\Omega = (\omega, A)$ of Ω_0 (that is, for every smooth map $p \mapsto \Omega(p)$ for which $\Omega(0) = \Omega_0$) defined on an open neighbourhood $\mathscr{P}$ of the origin of $\mathbb{R}^q$, the following holds. There is a smaller neighbourhood $\bar{\mathscr{P}} \subset \mathscr{P}$ of 0 and there are maps $\psi : \bar{\mathscr{P}} \to \Sigma$ and $C : \bar{\mathscr{P}} \to \mathsf{GL}(n,\mathbb{R})$, such that $\psi(0) = 0$, $C(0) = I$ and

$$\bar\omega(\psi(p)) = \omega(p) \qquad C(p)\bar A(\psi(p))C(p)^{-1} = A(p). \tag{7}$$

More generally, $\bar\Omega$ is a *versal unfolding* of Ω_0 in the Lie algebra $\mathbb{R}^m \times \mathfrak{g}$ of the Lie group $\mathbb{T}^m \times \mathfrak{G}$, if $\Omega_0 \in \mathbb{R}^m \times \mathfrak{g}$, and if for every smooth deformation Ω of Ω_0 taking values in $\mathbb{R}^m \times \mathfrak{g}$, maps $\psi : \mathscr{P} \to \Sigma$ and $C : \mathscr{P} \to \mathfrak{G}$ can be found such that the equations (7) hold.

The map Ω is called *miniversal* if the dimension of $\mathscr{P}$ is the smallest possible for a versal unfolding (see [1], §30).

2.1.7. *Diophanticity.* For any $k \in \mathbb{Z}^m$, let $|k| = \sum_{i=1}^m |k_i|$. Choose $\gamma_0, \kappa > 0$. A vector $\omega \in \mathbb{R}^m$ is called (γ_0,κ)–*Diophantine*, or *Diophantine* for short, if

$$|\langle k, \omega\rangle| \ge \gamma_0 |k|^{-\kappa}, \tag{8}$$

for all $k \in \mathbb{Z}^m \backslash \{0\}$. If $\kappa > m-1$ and if $\gamma_0 > 0$ is sufficiently small, the set of (γ_0,κ)–Diophantine vectors has positive Lebesgue measure in $\mathbb{R}^m$.

For $A \in \mathfrak{g}$, let $\alpha = \alpha_A$ be the vector of imaginary parts of the eigenvalues of A. If A depends continuously on a parameter p, the components of α are assumed to be arranged such that they depend continuously on p.

For $\gamma_0, \gamma, \kappa > 0$, the frequency $\Omega = (\omega, A)$ is said to be *normally* $(\gamma_0, \gamma, \kappa)$-*Diophantine*, or *normally Diophantine* for short, if ω is (γ_0, κ)-Diophantine, and if moreover $\alpha = \alpha_A$ satisfies

$$|\langle k, \omega\rangle + \langle \ell, \alpha\rangle| \geq \gamma(|k| + |\ell|)^{-\kappa},$$

for all $(k, \ell) \in \mathbb{Z}^m \times \mathbb{Z}^n$ such that $k \neq 0$ and $0 < |\ell| \leq 2$. This is indicated by writing $\Omega \in \mathrm{ND}_c = \mathrm{ND}_c(\gamma_0, \gamma, \kappa)$. Note that this definition does not depend on the arrangement of the components of α.

Let $\gamma_0, \gamma, \kappa > 0$ be fixed. A frequency map Ω is *quasi–periodically non–degenerate*, if Ω unfolds $\Omega(0)$ versally and $\Omega(0) \in \mathrm{ND}_c(\gamma_0, \gamma, \kappa)$. For Ω a given quasi–periodically nondegenerate frequency map, let

$$\mathscr{P}' = \{p \in \mathscr{P} \,:\, \Omega(p) \in \mathrm{ND}_c\}\,.$$

2.1.8. *Normal conjugacies.* The vector field X is said to be *normally conjugated* to a normal linear vector field L at a parameter value p, if there is a neighbourhood $\mathscr{U}$ of $\mathscr{T}$ and a conjugacy $\Phi(p) : \mathscr{U} \to \mathscr{M}$ such that

$$\mathsf{N}\Big(\Phi(p)_* X(p)\Big) = L(p).$$

Note that if L is tangent to $\mathscr{T}$, and X is normally conjugated to L, then X is tangent to the torus $\Phi^{-1}(\mathscr{T})$, and this torus is invariant under the flow of X.

Let $\pi_2 : \mathscr{M} \to \mathbb{R}^n$ be the projection $\pi_2(x, y) = y$. A conjugacy $\Phi : \mathscr{M} \to \mathscr{M}$ is said to be of *mean* τ if

$$\int_{\Phi^{-1}(\mathscr{T})} \pi_2 \cdot \Phi^* \,\mathrm{d}x = \int_{\mathbb{T}^m} (\pi_2 \circ \Phi^{-1})(x, 0)\,\mathrm{d}x = \tau.$$

2.2. **Differentiability classes.** The modifying terms theorem stated below will be proved for several differentiability classes.

2.2.1. *Notation.* Let $\mathscr{V} \subset \mathbb{R}^m$ be an open set, and let $\mathscr{W}$ a normed vector space. For a multi-index $\beta \in \mathbb{N}^m$, the β-derivative $D^\beta f$ with respect to $x \in \mathbb{R}^m$ of a $|\beta|$-times differentiable function f is defined as

$$D^\beta f = \frac{\partial^{|\beta|} f}{\partial x_1^{\beta_1} \cdots \partial x_m^{\beta_m}}.$$

2.2.2. *Finitely differentiable functions.* For $\mathscr{V}$ and $\mathscr{W}$ as above, let $f : \mathscr{V} \to \mathscr{W}$ be a continuous function that satisfies for some $0 < s < 1$ the inequality

$$|f(x) - f(y)| \leq C|x - y|^s, \qquad \text{for all } x, y \in \mathscr{V}.$$

Then f is *Hölder* continuous with exponent s. The smallest C such that the equality holds is the Hölder norm $\|f\|_s$ of f. The space of Hölder continuous functions $f : \mathscr{V} \to \mathscr{W}$ with Hölder exponent s is denoted by $\mathcal{C}^s = \mathcal{C}^s(\mathscr{V}, \mathscr{W})$. We write $\mathcal{C}^s(\mathscr{V})$ for $\mathcal{C}^s(\mathscr{V}, \mathbb{R})$.

Let $[s]$ denotes the largest integer smaller than or equal to $s \in \mathbb{R}$. For $s > 0$ and $s \neq \mathbb{N}$, an s–times differentiable function is an $[s]$ times continuously differentiable function $f : \mathscr{V} \to \mathscr{W}$, whose $[s]$–order partial derivatives $D^\beta f$ (where $|\beta| = [s]$) are Hölder continuous with exponent $s - [s]$ on $\mathscr{V}$. With the (recursive) definition of a norm

$$\|f\|_{\mathcal{C}^s} = \max_{|\beta| \leq [s]} \|D^\beta f\|_{\mathcal{C}^{s-[s]}},$$

the space of s–times differentiable functions is a Banach space, which will also be denoted by $\mathcal{C}^s$.

2.2.3. *Ultradifferentiable functions.* Let a sequence

$$M = \{M_j\}_{j=0}^{\infty}$$

be given, with $M_j > 1$ for every j. A smooth function $f : \mathscr{V} \to \mathscr{W}$ is said to be in the *Carleman class* $\mathcal{C}^M$ of ultradifferentiable functions, if there are constants $C, h > 0$, such that

$$\sup_{\mathscr{V}} |D^\beta f| \le C h^{-|\beta|} M_{|\beta|} \qquad \text{for every} \quad \beta \in \mathbb{N}^m. \tag{9}$$

Let $\|f\|_{\mathcal{C}_h^M}$ be the smallest constant C for which these estimates are satisfied: this defines the $\mathcal{C}_h^M$-norm of f, and with this norm $\mathcal{C}_h^M$ is a Banach space. Note that if $h_1 < h_2$, then $\mathcal{C}_{h_2}^M \subset \mathcal{C}_{h_1}^M$.

If $M_k = k!$, then $\mathcal{C}^M$ is the class $\mathcal{C}^\omega$ of real analytic functions, and $\mathcal{C}_h^M$ is the space $\mathcal{C}_h^\omega$ of real analytic functions that can be extended to complex analytic functions on a complex strip of width h in the imaginary direction. Since this class will be used extensively in the following, the norm $\|.\|_{\mathcal{C}_h^M}$ is written as $|.|_h$ in this case.

If $M_k = (k!)^\mu$, with $\mu > 1$, then $\mathcal{C}^M$ is the Gevrey class $\mathcal{G}^\mu$. The associated Gevrey spaces are denoted by $\mathcal{G}_h^\mu$. Unlike the real analytic class, for every $\mu > 1$ there are functions in $\mathcal{G}_h^\mu$ with compact support.

2.2.4. *Whitney smoothness.* The definitions of the function spaces just introduced can be extended to cover functies $f : \mathscr{F} \to \mathscr{W}$ that are defined on closed sets $\mathscr{F} \subset \mathscr{V}$, by replacing partial derivatives $D^\beta f$ with components f_β of a *Whitney jet* (cf. [34]). Let a collection of functions $\{f_\beta\}_\beta : \mathscr{F} \to \mathscr{W}$ be given such that $f_0 = f$ and such that the following consistency condition is satisfied for all β:

$$f_\beta(x+y) = \sum_{|\tilde\beta| \le s - |\beta|} f_{\beta+\tilde\beta}(x) \frac{y^{\tilde\beta}}{\tilde\beta!} + \mathsf{o}(|y|^{s-|\beta|}), \qquad x, y \in \mathscr{F}.$$

At every interior point x of $\mathscr{F}$ obviously $f_\beta(x) = D^\beta f(x)$. Finite differentiability and the smoothness classes $\mathcal{C}^M$ are now defined for functions on closed sets in the obvious way.

Whitney differentiable functions of a given smoothness class can be extended from $\mathscr{F}$ to all of $\mathscr{V}$; however, the results in this direction are increasingly weaker with increasing differentiability. For finite differentiability, there is a continuous linear extension operator [34]; for smooth functions, extension can still be shown to be a continuous operation [24]. Finally, Gevrey regular functions can be extended to Gevrey functions of the same class, but in general not continuously [5].

2.2.5. *Smoothness classes.* The regularity of conjugacies and invariant tori in the results below depends on the regularity of the data; to shorten the statement of the theorem, the following formalism is introduced: the original vector field and its perturbations (the "data") will be in a smoothness class $\mathcal{B}$, while mappings that are constructed in the proof in the theorem will be in a less regular class $\mathcal{B}'$, which depends on the original class $\mathcal{B}$. For each of the four $\mathcal{B}$-classes $\mathcal{C}^\omega$, $\mathcal{G}^\mu$, $\mathcal{C}^M$ and $\mathcal{C}^s$, we describe the corresponding $\mathcal{B}'$-class.

Let $\mathscr{U}$ be an open and bounded neighbourhood of $\mathscr{T} = \mathbb{T}^m \times \{0\}$. Functions in the $\mathcal{B}'$-classes are always more regular in the phase variables (x, y) than in the parameters p; we express this by positing that if $f \in \mathcal{B}'$, then for fixed values of p

$$f(.,.,p) \in \mathcal{B}_1'(\mathscr{U}, \mathscr{W})$$

and for fixed values of (x, y)

$$f(x, y, .) \in \mathcal{B}_2'(\mathscr{P}', \mathscr{W});$$

by specifying $\mathcal{B}_1'$ and $\mathcal{B}_2'$, we specify $\mathcal{B}'$. Note that since parameters are restricted to the closed set $\mathscr{P}'$, the smoothness of the parameter dependence is always meant in the sense of Whitney.

Let $\ell > 0$ be a positive integer, which denotes the maximal degeneracy of a normal eigenvalue of the unperturbed vector field X.

1. Analytic data. If $\mathcal{B} = \mathcal{C}^\omega_{2h}$, then for any $\zeta > 0$,

$$\mathcal{B}_1' = \mathcal{C}^\omega_h, \qquad \mathcal{B}_2' = \mathcal{G}^{1+\ell(\kappa+1)+\zeta}.$$

2. Gevrey regular data. If $\mathcal{B} = \mathcal{G}^\mu_h$ with $\mu > 1$, then for any $\zeta > 0$

$$\mathcal{B}_1' = \mathcal{G}^{\nu_1}_{h_1}, \qquad \mathcal{B}_2' = \mathcal{G}^{\nu_2}_{h_2},$$

where $h_1, h_2 > 0$ are some constants, and where

$$\nu_1 = 1 + \mu + \zeta, \qquad \nu_2 = 1 + \ell\mu(\kappa + 1) + \zeta.$$

3. Ultradifferentiable data. If $\mathcal{B} = \mathcal{C}^M_h(\mathscr{M} \times \mathscr{P})$, the description of $\mathcal{B}'$ is a little intricate.

If $f \in \mathcal{C}^M_h$, then for every multi-index α with $|\alpha| = s$ we have

$$\sup |D^\alpha f| \le C h^{-s} M_s.$$

Fix $\eta > 0$, and for every $s \in \mathbb{N}$ let $\lambda_s = (s + 1)\log C_0 + \eta s \log s + \log M_s$, where $C_0 = \max\{c_1/h, C\}$ with c_1 the constant given in lemma 4.1 below.

Let $\lambda_* : [0, \infty) \to \mathbb{R}$ be the largest convex function such that $\lambda_*(s) \le \lambda_s$ for $s \in \mathbb{N}$. Denote by $\mathcal{L}\lambda_*$ the Legendre transform of λ_*, which is given by

$$\mathcal{L}\lambda_*(p) = \max_{x\in[0,\infty)} \Big\{ px - \lambda_*(x) \Big\};$$

see subsection 4.3.2 below.

We construct a function g_M as follows. For a fixed constant $1 < \beta < 2$, chosen in the course of the proof, let $g_0 = \mathcal{L}\lambda_*(0)$ and let

$$g_j = \min\Big\{\beta g_{j-1}, \mathcal{L}\lambda_* \left(\log r_j^{-1}\right)\Big\}.$$

Finally, let g_M be the largest convex function such that

$$g_M(\log r_j^{-1}) \le g_j$$

for all j. Here $r_j = r_0 a_1^j$ with $0 < a_1 < 1$ and $r_0 > 0$, which are also chosen in the course of the proof.

Then $\mathcal{B}_1'$ and $\mathcal{B}_2'$ are respectively the Carleman classes $\mathcal{C}^{\tilde{M}^{(1)}}$ and $\mathcal{C}^{M^{(2)}}$, with

$$\tilde{M}^{(1)}_s = s!\, \mathrm{e}^{\mathcal{L}g_M(s+\tilde{C})}, \qquad \tilde{M}^{(2)}_s = s!\, \mathrm{e}^{\mathcal{L}g_M(\ell(\kappa+1+\zeta)s+\tilde{C})},$$

where $\zeta > 0$ is arbitrary, and where $\tilde{C}$ is a given constant.

4. Finitely differentiable data. Here $\mathcal{B} = \mathcal{C}^s$, with $s > (2n^2 + 3n)(\kappa + 1) + 3$. Then for any fixed $\zeta > 0$

$$\mathcal{B}_1' = \mathcal{C}^{s-(n^2+n)\kappa-2-\zeta}, \qquad \mathcal{B}_2' = \mathcal{C}^{(s-(n^2+n)\kappa-2-\zeta)/(\ell\kappa+\ell)}.$$

Note that always $\mathcal{B}' \subset \mathcal{C}^{n+1}$.

Remark 1. The conjugacies can be extended as maps, using the theorems mentioned above, to larger parameter sets; however, in general they will cease to be conjugacies on these larger sets.

Remark 2. The size of the open neighbourhood of the unperturbed vector field X for which the perturbation theorem below holds, will in general depend on the constant ζ.

2.2.6. *Vector fields.* Since all tangent bundles which will appear in this article are trivial, a vector field X is identified with its component map $F = (F_1, F_2) : \mathscr{M} \times \mathscr{P} \to \mathbb{R}^m \times \mathbb{R}^n$ by setting $X = F_1 \frac{\partial}{\partial x} + F_2 \frac{\partial}{\partial y}$. The classes of vertical vector fields whose components are of class $\mathcal{B}$ or $\mathcal{B}'$ are denoted by $\mathcal{X} = \mathcal{X}(\mathscr{M} \times \mathscr{P})$ or $\mathcal{X}' = \mathcal{X}'(\mathscr{U} \times \mathscr{P}')$, respectively. In particular, by $\mathcal{X}^\omega$, $\mathcal{X}^\mu$ $\mathcal{X}^M$ and $\mathcal{X}^s$ are respectively indicated the class of vector fields that are analytic, Gevrey regular, Carleman regular and finitely differentiable. The norms $\|X\|_{\mathcal{B}}$ of vector field X in $\mathcal{X}$ are defined analogously to the function norms above.

2.3. **Parametrised modifying terms theorem.** In order to formulate the main theorem, let the parameter space $\mathscr{P} \subset \mathbb{R}^m \times \mathfrak{g} \times \mathbb{R}^{\tilde{q}}$ be an open connected set. Write $p \in \mathscr{P}$ as $p = (\bar{\Omega}, \bar{p}) = (\bar{\omega}, \bar{A}, \bar{p})$, and let $(\mathfrak{g}, \mathfrak{h})$ define an admissible structure of vector fields.

Main Theorem. *Fix $\Omega_0 = (\omega_0, A_0) \in \mathrm{ND}_c$, and let a frequency map Ω be given as $\Omega(p) = \Omega_0 + \bar{\Omega}$. Let $X \in \mathfrak{h} \cap \mathcal{X}$ be an integrable vector field with normal linear part L_Ω. Then there exists an $\varepsilon_0 > 0$ such that for any perturbation $P \in \mathfrak{h} \cap \mathcal{X}$ with $\|P\|_{\mathcal{B}} < \gamma_0 \varepsilon_0$, the following holds.*

There is an integrable vector field $\Lambda \in \mathfrak{h} \cap \mathcal{X}'$, $\|\Lambda\|_{\mathcal{B}'} < C\|P\|_{\mathcal{B}}$, such that if $\Omega \in \mathrm{ND}_c$, then $X + P - \Lambda$ is normally conjugated to L_Ω by a vertical mean–0 structure-preserving conjugacy Φ in $\mathcal{B}'$. We have that Φ is normally linear in y and that $\|\Phi - \mathrm{id}\|_{\mathcal{B}'} < C\|P\|_{\mathcal{B}}$ for some $C > 0$.

The proof of this theorem is given in section 4.

3. **Persistence of tori.**

3.1. **Perturbations of non-linear integrable families.** We are interested in the following situation. Let $p \mapsto \Omega(p) = (\omega(p), A(p))$ be a frequency map, defined on a neighbourhood $\mathscr{P}$ of $0 \in \mathbb{R}^q$. Denote by $A_0 : \mathbb{R}^n \to \mathbb{R}^n$ the linear map given by the matrix $A(0)$; introduce $A_1(p) = A(p) - A(0)$, such that

$$A(p) = A_0 + A_1(p),$$

and such that $A_1(p) = \mathsf{O}(|p|)$.

Let $\mathscr{N}$ and $\mathscr{R}$ denote the kernel and the range of A_0, respectively. Choose complementary subspaces $\mathscr{N}^c$ and $\mathscr{R}^c$ to $\mathscr{N}$ and $\mathscr{R}$; that is,

$$\mathscr{N} + \mathscr{N}^c = \mathscr{R} + \mathscr{R}^c = \mathbb{R}^n, \qquad \mathscr{N} \cap \mathscr{N}^c = \mathscr{R} \cap \mathscr{R}^c = \{0\}.$$

Given these choices, there is a unique decomposition of a vector $z \in \mathbb{R}^n$ as a sum $z = z_1 + z_2$ with $z_1 \in \mathscr{R}$ and $z_2 \in \mathscr{R}^c$. Define projections $\pi_{\mathscr{R}}$ and $\pi^c_{\mathscr{R}}$ by setting $\pi_{\mathscr{R}} z = z_1$ and $\pi^c_{\mathscr{R}} z = z_2$; projections $\pi_{\mathscr{N}}$ and $\pi^c_{\mathscr{N}}$ are defined analogously.

Let $X \in \mathfrak{h} \cap \mathcal{X}$ be an integrable vector field of the form

$$X = \Big(\omega(p) + q_1(y, p)\Big) \frac{\partial}{\partial x} + \Big(A(p) y + q_2(y, p)\Big) \frac{\partial}{\partial y},$$

where $q_1 = \mathsf{O}(|y|)$ and $q_2 = \mathsf{O}(|y|^2)$. Note that we do not make any assumptions on the matrix $A(p)$ in terms of multiplicity or vanishing of eigenvalues, and that therefore the "standard" KAM theorem, as for instance in [9], is not applicable.

If for $\tau \in \mathbb{R}^n$ the torus $\mathscr{T}_\tau = \mathbb{T}^m \times \{y = \tau\}$ is invariant under X, we have necessarily that

$$A(p)\tau + q_2(\tau, p) = 0. \tag{10}$$

Introducing $\mu = \pi^c_{\mathscr{N}}\tau$ and $\nu = \pi_{\mathscr{N}}\tau$, and projecting equation (10) on both $\mathscr{R}$ and $\mathscr{R}^c$, we obtain

$$A_0\mu + \pi_{\mathscr{R}} A_1(p)(\mu+\nu) + \pi_{\mathscr{R}} q_2(\mu+\nu, p) = 0 \tag{11}$$

and

$$\pi^c_{\mathscr{R}}\left(A_1(p)(\mu+\nu) + q_2(\mu+\nu, p)\right) = 0. \tag{12}$$

Since $A_0 : \mathscr{N}^c \to \mathscr{R}$ is invertible and since $A_1(p) = \mathsf{O}(|p|)$, if p takes values in a neighbourhood of 0, then equation (11) can be solved for $\mu = \mu(\nu, p)$ as a function of ν and p. Let

$$\tau = \tau(\nu, p) = \mu(\nu, p) + \nu.$$

Substitution in equation (12) yields a function $f : \mathscr{N} \times \mathscr{P} \to \mathscr{R}^c$ such that if

$$0 = f(\nu, p) = \pi^c_{\mathscr{R}}\Big(A_1(p)\tau + q_2(\tau, p)\Big),$$

then the vector field X has an invariant torus at $y = \tau(\nu, p)$.

In the statement of the following theorem a map $F : \mathscr{V} \times \mathscr{P} \to \mathscr{W}$, where $\mathscr{V} \subset \mathbb{R}^n$ and where $\mathscr{W}$ is any finite dimensional vector space, is said to be $\mathcal{B}'$-smooth if $F(., p) \in \mathcal{B}'_1$ for fixed p and if $F(\tau, .) \in \mathcal{B}'_2$ for fixed τ.

Theorem 3.1. (Quasi-periodic Lyapunov-Schmidt reduction) *There exists an $\varepsilon_0 > 0$, independent of γ_0, such that for any $P \in \mathfrak{h} \cap \mathcal{X}$ with $\|P\|_{\mathcal{B}} < \gamma_0\varepsilon_0$ the following holds.*

There is a smaller neighbourhood $\bar{\mathscr{P}}$ of 0, a conjugacy $\Phi_p : \mathbb{T}^m \times \mathbb{R}^n \to \mathbb{T}^m \times \mathbb{R}^n$, a frequency map $\hat{\Omega} : \bar{\mathscr{P}} \to \mathbb{R}^m \times \mathfrak{g}$, and maps $\mu : \mathscr{N} \times \bar{\mathscr{P}} \to \mathscr{N}^c$, $\tau : \mathscr{N} \times \bar{\mathscr{P}} \to \mathbb{R}^n$, $f : \mathscr{N} \times \bar{\mathscr{P}} \to \mathscr{R}^c$, both $\mathcal{B}'$–smooth, as well as maps $\rho_1 : \mathbb{R}^n \times \bar{\mathscr{P}} \to \mathbb{R}^m$, $\rho_2 : \mathbb{R}^n \times \bar{\mathscr{P}} \to \mathbb{R}^n$, $\rho_3 : \mathbb{R}^n \times \bar{\mathscr{P}} \to \mathfrak{g}$, at least $\mathcal{C}^{n+1}$-smooth, with the following properties.

The map $\mu = \mu(\nu, p)$, with $\nu \in \mathscr{N}$ and $p \in \bar{P}$ solves the equation

$$0 = \pi_{\mathscr{R}}\Big[A(p)(\mu+\nu) + q_2(\mu+\nu, p) + \rho_2(\mu+\nu, p)\Big]. \tag{13}$$

The map $\tau = \tau(\nu, p)$ is of the form

$$\tau(\nu, p) = \mu(\nu, p) + \nu. \tag{14}$$

The map f is of the form

$$f(\nu, p) = \pi^c_{\mathscr{R}}\Big[A(p)\tau + q_2(\tau, p) + \rho_2(\tau, p)\Big]. \tag{15}$$

The frequency map $\hat{\Omega}$ reads as

$$\hat{\Omega}(\nu, p) = \left(\omega(p) + q_1(\tau, p) + \rho_1(\tau, p), A(p) + \frac{\partial q_2}{\partial y}(\tau, p) + \rho_3(\tau, p)\right).$$

Moreover

$$\|\Phi - \mathrm{id}\|_{\mathcal{B}'} \to 0 \quad \textit{and} \quad \|\rho_i\|_{\mathcal{C}^{n+1}} \to 0$$

as $\|P\|_{\mathcal{B}} \to 0$.

Finally, if $\tau = \tau(\nu, p)$ and

$$f(\nu, p) = 0,$$

then Φ is a mean-τ conjugacy that normally conjugates $X + P$ to $L_{\hat{\Omega}}$ at all parameters for which $\hat{\Omega}(p)$ is normally Diophantine.

Proof. The proof runs along the same lines as the example of the introduction. Write $Q = X - \mathsf{N}X = q_1 \frac{\partial}{\partial x} + q_2 \frac{\partial}{\partial y}$ and set

$$\tilde{X} = X + P = L_\Omega + Q + P.$$

Let $\Psi_\tau : \mathscr{M} \times \mathscr{P} \to \mathscr{M} \times \mathscr{P}$ be a *localising transformation*, given by

$$\Psi_\tau^{-1}(x, y, p) = (x, \tau + y, p)\,.$$

Introduce the localised vector field

$$Y = \Psi_{\tau *} \tilde{X}.$$

Its normal linear part takes the form

$$\begin{aligned} \mathsf{N}Y = &\Big(\omega(p) + q_1(\tau, p)\Big) \frac{\partial}{\partial x} \\ &+ \left(A(p)\tau + q_2(\tau, p) + \left(A(p) + \frac{\partial q_2}{\partial y}(\tau, p) \right) y \right) \frac{\partial}{\partial y} \\ &+ \mathsf{N}\Psi_{\tau *} P. \end{aligned}$$

Let $\hat{\Omega} = (\hat{\omega}, \hat{A}) \subset \mathbb{R}^m \times \mathfrak{g}$, and introduce

$$\begin{aligned} \Lambda_0 = &\Big(\omega(p) + q_1(\tau, p) - \hat{\omega}\Big) \frac{\partial}{\partial x} \\ &+ \left(A(p)\tau + q_2(\tau, p) + \left(A(p) + \frac{\partial q_2}{\partial y}(\tau, p) - \hat{A} \right) y \right) \frac{\partial}{\partial y} \end{aligned}$$

Then

$$\mathsf{N}(Y - \Lambda_0) = L_{\hat{\Omega}} + \mathsf{N}\Psi_{\tau *} P.$$

Applying theorem 2.3 to $Y - \Lambda_0$ yields that for $\|P\|_{\mathcal{B}} = \varepsilon$ sufficiently small, there is a $\mathcal{B}'$–smooth integrable vector field

$$\Lambda_1 = \delta(p, \tau, \hat{\Omega}) \frac{\partial}{\partial x} + (\mu(p, \tau, \hat{\Omega}) + B(p, \tau, \hat{\Omega}) y) \frac{\partial}{\partial y},$$

such that $\|\Lambda_1\|_{\mathcal{B}'} \leq C\varepsilon$, and a $\mathcal{B}'$-smooth conjugacy

$$\Phi^{-1}(x, y; p, \tau, \hat{\Omega}) = \Big(x + \varphi_1(x, y, p, \tau, \hat{\Omega}), y + \varphi_2(x, y, p, \tau, \hat{\Omega}) \Big)\,,$$

such that

$$\mathsf{N}\Phi_*(Y - \Lambda_0 - \Lambda_1) = L_{\hat{\Omega}}. \tag{16}$$

The modifying terms vector field Λ_1 can be extended, non-uniquely, to a vector field defined for all $\hat{\Omega}$ that is at least $\mathcal{C}^{n+1}$, see [34], and which will also be denoted by Λ_1.

Putting $\Lambda = \Lambda_0 + \Lambda_1$, we want to determine a map $\hat{\Omega} : \bar{\mathscr{P}} \times \mathbb{R}^n \to \mathbb{R}^m \times \mathfrak{g}$ such that if $\hat{\Omega} = \hat{\Omega}(p, \tau)$, then $\Lambda = 0$. Requiring that $\Lambda = 0$ is equivalent to the equations

$$\hat{\omega} = \omega(p) + q_1(\tau, p) + \hat{\delta}(p, \hat{\Omega}), \tag{17}$$

$$0 = A(p)\tau + q_2(\tau, p) + \hat{\mu}(p, \hat{\Omega}), \tag{18}$$

$$\hat{A} = A(p) + \frac{\partial q_2}{\partial y}(\tau, p) + \hat{B}(p, \hat{\Omega}). \tag{19}$$

Since the modifying terms are $(\mathcal{C}^{n+1}, C\varepsilon)$-small, if $\varepsilon > 0$ is sufficiently small, equations (17) and (19) can be solved for $\hat{\omega}$ and $\hat{A}$, yielding

$$\hat{\omega}(\tau, p) = \omega(p) + q_1(\tau, p) + \rho_1(\tau, p),$$
$$\hat{A}(\tau, p) = A(p) + \frac{\partial q_2}{\partial y}(\tau, p) + \rho_3(\tau, p).$$

For $\mu \in \mathscr{N}^c$ and $\nu \in \mathscr{N}$, set $\tau = \mu + \nu$. By applying $\pi_{\mathscr{R}}$ to both sides of equation (18), and recalling that $A(p) = A_0 + A_1(p)$ with $A_1(p) = \mathsf{O}(|p|)$, we obtain

$$0 = A_0\mu + \pi_{\mathscr{R}}\left(A_1(p)(\mu + \nu) + q_2(\mu + \nu, p) + \rho_2(\mu + \nu, p)\right). \tag{20}$$

Note that $A_0 : \mathscr{N}^c \to \mathscr{R}$ is invertible. By the implicit function theorem, we can solve equation (14) for $\mu = \mu(\nu, p)$. Substituting this function into equation (18) and consequently applying $\pi^c_{\mathscr{R}}$ to both sides yields, with $\tau = \mu(\nu, p) + \nu$:

$$0 = f(\nu, p) \stackrel{\text{def}}{=} \pi^c_{\mathscr{R}}\Big[A(p)\tau + r(\tau, p) + \rho_2(\tau, p)\Big].$$

□

3.2. **Corollaries.** Note that in the situation of theorem 3.1, the linear map A_0 is invertible, then $\dim \mathscr{R}^c = 0$, and the equation $f = 0$ disappears. Moreover, if Ω is a versal unfolding of $\Omega(0)$, then so is $\hat{\Omega}$, and the set of parameters p such that $\hat{\Omega}(p) \in \mathrm{ND}_c$ has positive Lebesgue measure.

3.3. **Reduction of parameters.** The previous results can also be applied to situations with few parameters. The reduction is based on the following result of Pyartli.

Theorem 3.2. (Pyartli [29]). *Let U be an open neighbourhood of a point $q \in \mathbb{R}^m$, and let a smooth map $\alpha : \mathbb{R}^m \to \mathbb{R}^n$ $(n > m)$ be given, parametrising a m–dimensional submanifold S in $\mathbb{R}^n$. Assume that there is a curve $\xi : (-\varepsilon, \varepsilon) \to \mathbb{R}^m$ with $\xi(0) = q$, such that $v_1, \cdots, v_{n-m+1}$ span a $(n-m+1)$–dimensional linear subspace of $T_{\alpha(q)}\mathbb{R}^n$ transversal to $T_{\alpha(q)}S$ at $\alpha(q)$, where v_j is given as*

$$v_j = \frac{\mathrm{d}^j \alpha \circ \xi}{\mathrm{d}t^j}(0).$$

If $\kappa > n^2 - n + 1$, and if $\gamma > 0$ is sufficiently small, then the set

$$U_c = \left\{x \in U \,:\, |\langle k, \alpha(x)\rangle + k_0| \geq \gamma |k|^{-\kappa} \text{ for all } k \in \mathbb{Z}^m \backslash \{0\}, k_0 \in \mathbb{Z}\right\}.$$

has positive Lebesgue measure in U.

The significance of this theorem is expressed by the following, less precise, reformulation: if $\kappa > 0$ is sufficiently large, then for a generic frequency map Ω, the inverse image $\Omega^{-1}(\mathrm{ND}_c)$ has positive Lebesgue measure.

Suppose Ω is a frequency map such that $\Omega(0) \in \mathrm{ND}_c$. It is always possible to find a versal unfolding $\hat{\Omega}$ of $\Omega(0)$, defined on another parameter space Σ, such that Ω is a subfamily of $\hat{\Omega}$; that is, such that there is a map $\sigma : \mathscr{P} \to \Sigma$ with the property that

$$\Omega(p) = \hat{\Omega}(\sigma(p)).$$

Hence, a given vector field $X(p) = L_\Omega(p) + Q(p)$ – only the parameter dependence is made explicit – can be replaced by $\hat{X}(p, \sigma) = L_{\hat{\Omega}}(\sigma) + Q(p)$ with $(p, \sigma) \in \mathscr{P} \times \Sigma$. By theorem 2.3, for every small perturbation $P(p)$ there is an integrable vector field $\hat{\delta}(p, \sigma)$ such that $\hat{X} + P + \hat{\delta}$ has an invariant quasi–periodic torus of mean 0 whenever $\hat{\Omega}(\sigma) \in \mathrm{ND}_c$, since $\hat{\Omega}$ is quasi–periodically nondegenerate. Then, by using that

$$X(p) = \hat{X}(p, \sigma(p)),$$

the conclusion is obtained that for a generic set of vector fields X, there is an integrable vector field $\delta(p) = \hat{\delta}(p, \sigma(p))$, such that the set of parameters p for which $X(p) + P(p) + \delta(p)$ has an invariant quasi–periodic torus of mean 0, has positive Lebesgue measure in $\mathscr{P}$.

3.4. **The quasi-periodic Bogdanov-Takens bifurcation.** As an application of theorem 3.1, we treat the persistence of invariant tori in the quasi-periodic Bogdanov-Takens bifurcation [35, 3, 4, 10, 30, 11].

3.4.1. *Integrable normal form.* Recall that a Bogdanov-Takens singularity occurs if a singular point, say $x = 0$, of a planar vector field Z_0, has a multiple eigenvalue 0 with geometric multiplicity 1; that is, the linearisation has a nilpotent part. We assume that Z_0 is a member of a family of vector fields Z_σ, parametrised by a two-dimensional parameter σ. If some nondegeneracy conditions are met, by a suitable change of phase space and parameter space coordinates, the vector field can be brougth into the form

$$Z_\sigma = \left(-\begin{pmatrix} 0 \\ \sigma_1 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & \sigma_2 \end{pmatrix} y + \begin{pmatrix} 0 \\ y_1^2 + b y_1 y_2 \end{pmatrix} + r(y, \sigma) \right) \frac{\partial}{\partial y},$$

where $b = \pm 1$ and $r = \mathsf{O}(|y|^3)$. Note that Z_σ is an unfolding of the nilpotent singularity $y = 0$. We shall limit our attention to the case $b = 1$.

Consider now the integrable unfolding X_σ of the normally nilpotent invariant torus $\mathscr{T}_0 = \{(x, y) \in \mathbb{T}^m \times \mathbb{R}^2\}$ of the vector field X_0, where

$$X_\sigma = \omega(y, \sigma) \frac{\partial}{\partial x} + \left(-\begin{pmatrix} 0 \\ \sigma_1 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & \sigma_2 \end{pmatrix} y + \begin{pmatrix} 0 \\ y_1^2 + y_1 y_2 \end{pmatrix} + r(y, \sigma) \right) \frac{\partial}{\partial y},$$

Introduce the standard basis vectors $e_1 = (1, 0)$ and $e_2 = (0, 1)$. In terms of subsection 3.1, we have

$$A(\sigma) = \begin{pmatrix} 0 & 1 \\ 0 & \sigma_2 \end{pmatrix}, \quad \ker A(0) = \mathscr{N} = \mathbb{R} e_1 \quad \operatorname{ran} A(0) = \mathscr{R} = \mathbb{R} e_1.$$

We choose

$$\mathscr{N}^c = \mathscr{R}^c = \mathbb{R} e_2.$$

Let π_1 and π_2 be the projections on $\mathbb{R}e_1$ and $\mathbb{R}e_2$ respectively. Then $\pi_{\mathscr{N}} = \pi_{\mathscr{R}} = \pi_1$ and $\pi^c_{\mathscr{N}} = \pi^c_{\mathscr{R}} = \pi_2$.

3.4.2. *Non-integrable perturbation.* Consider a non-integrable perturbation $X_\sigma + P_\sigma$ of X_σ, where the perturbation term P_σ is such that $\|P_\sigma\|_{\mathcal{B}} < \varepsilon$.

We shall assume that the smoothness class $\mathcal{B}$ contains $\mathcal{C}^s$, where $s > 0$ is such that $\mathcal{B}'$ contains at least $\mathcal{C}^4$. For sufficiently small $\varepsilon > 0$, theorem 3.1 ensures the existence of a $\mathcal{B}'$-smooth map Φ and functions μ, τ, f, ρ_1, ρ_2, ρ_3, such that $\|\rho_i\|_{\mathcal{C}^3} \leq C\varepsilon$, and such that the following hold.

Writing $\mu = (0, \tau_2)$, $\nu = (\tau_1, 0)$ and $\tau = \mu + \nu$, the function $\tau_2 = \tau_2(\tau_1, p)$ solves equation (13), which takes the form

$$\begin{aligned} 0 &= \pi_1 \left[-\begin{pmatrix} 0 \\ \sigma_1 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix} + \begin{pmatrix} 0 \\ \tau_1^2 + \tau_1 \tau_2 \end{pmatrix} + r + \rho_2 \right] \\ &= \tau_2 + \pi_1 \left[r(\tau, \sigma) + \rho_2(\tau, \sigma) \right]. \end{aligned}$$

We find that

$$\tau_2 = 0 + \tau_1^3\eta_1(\tau_1,\sigma) + \varphi_1(\tau_1,\sigma),$$

where $\eta_1 \in \mathcal{C}^{s-3}$ and $\|\varphi_1\|_{\mathcal{C}^3} \le C\varepsilon$. Substitution into equation (15) yields the function f, which is of the form

$$\begin{aligned} f(\tau_1,\sigma) &= \pi_2\left[-\begin{pmatrix}0\\ \sigma_1\end{pmatrix} + \begin{pmatrix}0 & 1\\ 0 & \sigma_2\end{pmatrix}\tau + \begin{pmatrix}0\\ y_1^2 + y_1y_2\end{pmatrix} + r + \rho_2\right] \\ &= -\sigma_1 + \tau_1^2 + \pi_2(r+\rho_2). \end{aligned}$$

We find

$$f(\tau_1,\sigma) = \tau_1^2 - \sigma_1 + \tau_1^3\eta_2 + \varphi_2,$$

where $\eta_2 \in \mathcal{C}^{s-3}$ and $\|\varphi_2\|_{\mathcal{C}^3} \le C\varepsilon$. The frequency map takes the form

$$\begin{aligned} \hat\Omega(\tau_1,\sigma) &= \Big(\hat\omega(\tau_1,\sigma), \hat A(\tau_1,\sigma)\Big) \\ &= \left(\omega(\tau_1,0,\sigma), \begin{pmatrix}0 & 1\\ 2\tau_1 & \sigma_2+\tau_1\end{pmatrix}\right) + \tau_1^2\eta_3 + \varphi_3, \end{aligned}$$

with $\eta_3 \in \mathcal{C}^{s-4}$ and $\|\varphi_3\|_{\mathcal{C}^3} \le C\varepsilon$. We solve σ_1 from the equation $f = 0$ to obtain

$$\sigma_1 = \Sigma(\tau_1,\sigma_2) = \tau_1^2 + \tau_1^3\eta_4(\tau_1,\sigma_2) + \varphi_4(\tau_1,\sigma_2), \tag{21}$$

with $\eta_4 \in \mathcal{C}^{s-3}$ and $\|\varphi_4\|_{\mathcal{C}^3} \le C\varepsilon$.

Theorem 3.1 then allows us to conclude that if $\sigma_1 = \Sigma(\tau_1,\sigma_2)$ and if $\hat\Omega$ is normally Diophantine, then $X_\sigma + P_\sigma$ has an invariant m-dimensional torus that is of the form $\mathscr{T}_\sigma = \{(x,y) : y = (\tau_1,\tau_2) + \varphi(x,y,\sigma)\}$, where $\|\varphi\|_{\mathcal{B}'} \le C\varepsilon$, with normal dynamics $L_{\hat\Omega}$.

3.4.3. *Quasi-periodic saddle-node bifurcations of $X_\sigma + P_\sigma$.* Let σ_1^* be a critical value of the map

$$(\tau_1,\sigma_2) \mapsto \Sigma(\tau_1,\sigma_2),$$

corresponding to a critical point (τ_1^*,σ_2^*). Write $\sigma^* = (\sigma_1^*,\sigma_2^*)$. If $\hat\Omega(\tau_1^*,\sigma^*)$ is normally Diophantine, then σ^* is a quasi-periodic saddle-node bifurcation point. It follows from (21) that the critical points of Σ satisfy

$$\tau_1 = 0 + \varphi_5(\sigma_2),$$

where $\|\varphi_5\|_{\mathcal{C}^1} \le C\varepsilon$.

3.4.4. *Quasi-periodic Hopf bifurcations.* At parameters for which the normal frequencies of an invariant m-torus are located on the imaginary axis, quasi-periodic Hopf bifurcations can occur. The full normal form analysis is not given here, but it runs along entirely standard lines. From the normal part $\hat A$ of the frequency map, we obtain the conditions

$$T(\tau_1,\sigma) = \mathrm{tr}\hat A = \tau_1 + \sigma_2 + \tau_1^2\eta_6 + \varphi_6 = 0 \tag{22}$$

and

$$D(\tau_1,\sigma) = \det \hat A = -2\tau_1 + \tau_1^2\eta_7 + \varphi_7 > 0.$$

Note that necessarily at all quasi-periodic saddle-node bifurcation points σ^*, with corresponding $\tau_1^* = \varphi_5(\sigma_2^*)$, we have

$$D(\tau_1^*,\sigma^*) = 0.$$

Solving equation (22) for τ_1, we obtain

$$\tau_1 = -\sigma_2 + \varphi_8(\sigma);$$

substitution in (21) yields the locus of the quasi-periodic Hopf bifurcation points as those parameter values σ such $\hat{\Omega}(-\sigma_2 + \varphi_8, \sigma)$ in normally Diophantine, for which

$$\sigma_1 = \Sigma(-\sigma_2 + \varphi_8(\sigma), \sigma_2),$$

as long as $D(-\sigma_2 + \varphi_8, \sigma) > 0$. This yields

$$\sigma_1 = \sigma_2^2 + \sigma_2^3 \eta_9 + \varphi_9.$$

These bifurcation curves are illustrated in figure 1.

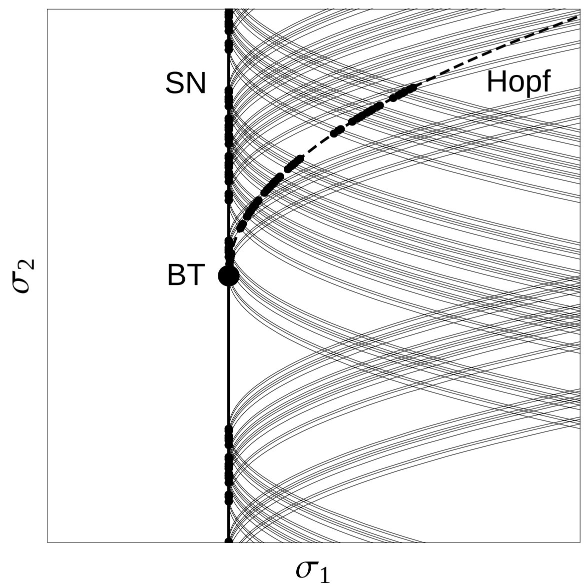

FIGURE 1. Bifurcation diagram of the quasi-periodic Bogdanov-Takens bifurcation

4. **Proof of the main result.**

In this section the proof of theorem 2.3 is given.

4.1. **Preliminaries.**

The vector field X mentioned in the statement of theorem 2.3 is defined on the phase space $\mathbb{T}^m \times \mathbb{R}^n$, it is integrable, and it has normal linear part L_Ω, where $\Omega = (\omega, A)$. Hence, it is of the form

$$X = (\omega + q_1(y,p)) \frac{\partial}{\partial x} + (Ay + q_2(y,p)) \frac{\partial}{\partial y},$$

with $q_1 = \mathsf{O}(|y|)$ and $q_2 = \mathsf{O}(|y|^2)$. The perturbation term P will be written as

$$P = p_1(x,y,p)\frac{\partial}{\partial x} + p_2(x,y,p)\frac{\partial}{\partial y};$$

it satisfies $\|P\|_{\mathcal{B}} \leq \gamma_0 \varepsilon_0$. In the following the vector field $X + P$ shall be denoted by $\tilde{X}$. After scaling the time by $t = \gamma_0 t'$, it may be assumed that the Diophantine condition ND_c is of the form $\mathrm{ND}_c(1, \gamma/\gamma_0, \kappa)$, and that $\|P\|_{\mathcal{B}} = \varepsilon < \varepsilon_0$.

Note that the frequency map $\Omega(p) = \Omega_0 + \bar{\Omega}$ is a linear function of p.

4.1.1. *Multiple normal eigenvalues.* For the following remarks, cf. [20, 7]. In order to motivate the definition of the parameter domains below, we need some estimates on the parameter dependence of eigenvalues in the case that the matrix A_0 has multiple eigenvalues.

Let f be the characteristic polynomial of $A(p) = A_0 + \bar{A}$; that is,

$$f(z,p) = \det(A(p) - zI).$$

If $\lambda \in \mathbb{C}$ is an ℓ-fold zero of $f(z,0)$, then by the Weierstraß preparation theorem (see for instance [21], p. 155), there are unique analytic functions $q(z,p)$, $a_i(p)$, defined in a neighbourhood of $(z,p) = (\lambda, 0)$, such that $q(\lambda,0) \neq 0$, $a_i(0) = 0$ for $i = 0, \cdots, n-1$, and

$$(z-\lambda)^\ell = qf + \sum_{i=0}^{\ell-1} a_i(p)(z-\lambda)^i.$$

The function $g(z,p) = 1/q(z,p)$ is defined in a, possibly smaller, neighbourhood of $(\lambda,0)$, and

$$f(z,p) = g(z,p)\left((z-\lambda)^\ell - \sum_{i=0}^{\ell-1} a_i(p)(z-\lambda)^i\right).$$

There are ℓ continuous functions $z_k(p)$, $k = 1, \cdots, \ell$, defined for p in an open bounded neighbourhood U of 0, such that $z_k(0) = \lambda$ and such that

$$\prod_{i=1}^{\ell}(z - z_k(p)) = (z-\lambda)^\ell - \sum_{i=0}^{\ell-1} a_i(p)(z-\lambda)^i.$$

For $z \in \mathbb{C}^N$, introduce the norm

$$|z| = \max_{0\le i\le N} |z_i|. \tag{23}$$

The functions z_k satisfy

$$|z_k(p)| < C|p|^{1/\ell} \tag{24}$$

for some $C > 0$. To see this, assume (as we may) that U is the common domain of definition for the functions $a_i(p)$ and $z_k(p)$. Since the $a_i(p)$ are analytic and satisfy $a_i(0) = 0$, there is a constant $C' > 0$ such that $|a_i(p)| < C'|p|$ on U. For $|p| < 1/(\ell C')$ and $|z| \ge 1$, it follows that

$$|f_\lambda(z,p)| > |z|^\ell - \sum_{i=0}^{\ell-1} |a_i(p)||z|^i > 0;$$

consequently $|z_k(p)| < 1$ if $|p| < 1/(\ell C')$, and then $f_\lambda(z_k(p),p) = 0$ implies that

$$|z_k(p)|^\ell \le \sum_{i=0}^{\ell-1} |a_i(p)| < \ell C'|p|.$$

In turn, this implies inequality (24). We conclude that the eigenvalues of $A(p)$ are Hölder continuous as a function of p. The Hölder exponent is equal to $1/\ell$, where ℓ is the largest multiplicity of an eigenvalue of A_0. Moreover, for all p such that the eigenvalues of $A(p)$ are all different, they depend analytically on p.

4.1.2. *Parameter domains.* Define the distance d between two points $x, y \in \mathbb{C}^N$ as $d(x, y) = |x - y|$, where the norm $|.|$ has been introduced in equation (23). If $\mathscr{U} \subset \mathbb{C}^N$, the distance of a point x to $\mathscr{U}$ is given as $d(x, \mathscr{U}) = \inf_{y \in \mathscr{U}} d(x, y)$. Define the open complex strip $\mathscr{U} + r$ of width r around a set $\mathscr{U}$ by

$$\mathscr{U} + r = \left\{ z \in \mathbb{C}^N \,:\, d(z, \mathscr{U}) < r \right\}.$$

Let $\{d_j\}$ be a given sequence of positive real numbers, monotonically increasing towards infinity. Let the set $\mathrm{nd}_c^j \subset \mathbb{R}^m \times \mathbb{R}^n$ of normally Diophantine frequencies be the set of vectors (ω, α), where α is of the form

$$\alpha = (\alpha_1, -\alpha_1, \cdots, \alpha_k, -\alpha_k, 0, \cdots, 0),$$

such that the conditions

$$|\langle k, \omega \rangle| \geq \gamma_0 |k|^{-\kappa}, \qquad |\langle k, \omega \rangle + \langle \ell, \alpha_A \rangle| \geq \gamma \left(|k| + |\ell|\right)^{-\kappa},$$

are satisfied for all $(k, \ell) \in \mathbb{Z}^m \times \mathbb{Z}^n$ for which $0 < |k| \leq d_j$, $|\ell| \leq 2$.

For given $\Omega = (\omega, A) \in \mathbb{T}^m \times \mathfrak{g}$, let α_A be the vector of imaginary parts of eigenvalues of A. Introduce the set $\mathrm{ND}_c^j \subset \mathbb{T}^m \times \mathfrak{g}$ of normally Diophantine $\Omega = (\omega, A)$ by requiring that their frequency vectors (ω, α_A) are normally Diophantine.

Furthermore, if $\{\rho_j\}$ is a positive sequence that decreases monotonically to 0, let $\mathrm{ND}_c^j(\rho_j) \subset \mathbb{T}^m \times \mathfrak{g}$ be the set of $\Omega = (\omega, A)$ such that their frequency vector (ω, α_A) satisfies $|\omega - \bar{\omega}| + |\alpha_A - \bar{\alpha}| < \rho_j$, where $(\bar{\omega}, \bar{\alpha}) \in \mathrm{nd}_c^j$.

Note that

$$\mathrm{ND}_c^{j+1} \subset \mathrm{ND}_c^j, \qquad \mathrm{ND}_c^{j+1}(\rho_{j+1}) \subset \mathrm{ND}_c^j(\rho_j),$$

and that

$$\bigcap_{j=1}^{\infty} \mathrm{ND}_c^j = \bigcap_{j=1}^{\infty} \mathrm{ND}_c^j(\rho_j) = \mathrm{ND}_c.$$

Finally, introduce

$$\mathscr{P}(\rho_j) = \left\{ p \in \mathscr{P} \,|\, \Omega(p) \in \mathrm{ND}_c^j(\rho_j) \right\},$$

and note that $\mathscr{P}(\rho_{j+1}) \subset \mathscr{P}(\rho_j)$ and $\cap_{j=1}^{\infty} \mathscr{P}(\rho_j) = \mathscr{P}'$.

Take $p \in \mathscr{P}'$ and $\tilde{p} \in \mathscr{P} \backslash \mathscr{P}(\rho_j)$. Recall from subsubsection 4.1.1 that the normal eigenvalues $\alpha_A(p)$ are Hölder continuous with Hölder exponent ℓ, where ℓ is the highest algebraic multiplicity of an eigenvalue of A_0. Then

$$\rho_j \leq |\omega(\tilde{p}) - \omega(p)| + |\alpha_A(\tilde{p}) - \alpha_A(p)| < C |\tilde{p} - p|^{1/\ell},$$

and $|\tilde{p} - p| > \rho_j^\ell / C^\ell$. As a consequence, we have that

$$\mathscr{P}' + \frac{1}{C^\ell} \rho_j^\ell \subset \mathscr{P}(\rho_j). \tag{25}$$

4.1.3. *Phase domains.* Let $\mathscr{U}$ be an open subset of $\mathscr{M}$. An embedding $\Phi : \mathscr{U} \times \mathscr{P}$ is called vertical over the parameters, if it acts as the identity on the space of parameters, that is, if it can be written in the form $\Phi(x, y, p) = (\Phi_p(x, y), p)$. In the course of the proof, a sequence $\{\Phi_j\}$ of vertical embeddings will be constructed inductively, together with infinite sequences of complex domains $\{\mathscr{D}_j\}$ and $\{\tilde{\mathscr{D}}_j\}$. The definitions of the domains are slightly different according to whether $\tilde{X}$ is in the real analytic class or not.

Let $\{r_j\}, \{\rho_j\}, \{\tilde{r}_j\}, \{\tilde{\rho}_j\}$ be geometrically decreasing sequences, which will be chosen later on, but which are from the outset assumed to satisfy $r_1 \leq \tilde{r}_1$ and $\rho_1 \leq \tilde{\rho}_1$.

Let $\mathscr{V}$ be an open bounded real neighbourhood of $\mathscr{T} = \mathbb{T}^m \times \{0\}$. If $\tilde{X}$ is real analytic, then there is some constant $h > 0$ such that $\tilde{X}$ can be extended to an analytic vector field on $\mathscr{V} + 2h$, which is an open neighbourhood of $\mathscr{T}$ in $\mathbb{T}^m_{\mathbb{C}} \times \mathbb{C}^n$ (where $\mathbb{T}^m_{\mathbb{C}} = \mathbb{C}^m/2\pi\mathbb{Z}^m$). Let in this case $\mathscr{U}$ be the complex neighbourhood $\mathscr{V} + h$ of $\mathscr{V}$; otherwise, if no analytic extension of $\tilde{X}$ to a complex neighbourhood of $\mathscr{V}$ exists, let $\mathscr{U}$ be equal to $\mathscr{V}$.

The domains $\mathscr{D}_j$ and $\tilde{\mathscr{D}}_j$ are defined in terms of $\mathscr{U}$ as follows

$$\mathscr{D}_j = \mathscr{D}(r_j, \rho_j) = (\mathscr{U} + r_j) \times \mathscr{P}(\rho_j), \tag{26}$$

$$\tilde{\mathscr{D}}_j = \tilde{\mathscr{D}}(\tilde{r}_j, \tilde{\rho}_j) = ((\Phi_j)_p(\mathscr{U}) + \tilde{r}_j) \times \mathscr{P}(\tilde{\rho}_j). \tag{27}$$

In the following, also "intermediate" domains $\mathscr{D}_{j+1} \subset \mathscr{D}_{j+\vartheta} \subset \mathscr{D}_j$ are needed. For $0 < \vartheta < 1$, define first the convex combinations $r_{j+\vartheta} = \vartheta r_{j+1} + (1-\vartheta) r_j$ and $\rho_{j+\vartheta} = \vartheta \rho_{j+1} + (1-\vartheta)\rho_j$, and then

$$\mathscr{D}_{j+\vartheta} = \mathscr{D}(r_{j+\vartheta}, \rho_{j+\vartheta}).$$

For analytic functions on some complex open set $\mathscr{O}$, the norm

$$|f|_{\mathscr{O}} = \sup_{\mathscr{O}} |f(x, y, p)|$$

is introduced; if $\mathscr{O} = (\mathscr{U} + \sigma_1) \times \mathscr{P}(\sigma_2)$ this is abbreviated to $|f|_\sigma$; if $\mathscr{O} = \mathscr{D}_{j+\vartheta}$, it is further abbreviated to $|f|_{j+\vartheta}$.

4.2. **Structure of the proof.** One of the main technical problems of the proof is to deal with the smoothness of the vector field $\tilde{X}$ in the non–analytic cases. We shall work with analytic approximations: in the first part of the proof a sequence of analytic vector fields $\{\tilde{X}_j\}$ is constructed, where $\tilde{X}_j$ is defined on $\tilde{\mathscr{D}}_j$, which tends to $\tilde{X}$ in an appropriate sense.

In the second part of the proof, coordinate transformations Φ_j, "modifying terms" vector fields Λ_j and auxiliary vector fields X_j, Δ_j and $\tilde{\Delta}_j$ are constructed inductively by the following "staircase construction".

To set up the induction, choose

$$\Phi_1 : \mathscr{D}_1 \hookrightarrow \tilde{\mathscr{D}}_1$$

as the identity $(\Phi_1)_p(x, y) = (x, y)$, $\Lambda_0 = 0$, and X_1 as the restriction of $\tilde{X}_1$ to $\mathscr{D}_1$. Note that due to the assumptions $r_1 \le \tilde{r}_1$ and $\rho_1 \le \tilde{\rho}_1$, we have that $\mathscr{D}_1 \subset \tilde{\mathscr{D}}_1$, so that Φ_1 is well-defined.

At the beginning of the induction step, assume that an embedding

$$\Phi_j : \mathscr{D}_j \hookrightarrow \tilde{\mathscr{D}}_1,$$

a domain $\tilde{\mathscr{D}}_j$ of the form (27), and an integrable vector field Λ_j defined on $\tilde{\mathscr{D}}_j$ and another vector field X_j defined on $\mathscr{D}_j$ are already determined.

During the induction step, an embedding

$$\Psi_j : \mathscr{D}_{j+1} \hookrightarrow \mathscr{D}_j$$

and vector fields Δ_j on $\mathscr{D}_{j+\frac{1}{2}}$, and $\tilde{\Delta}_j$ and Λ_{j+1} on $\Phi_j(\mathscr{D}_{j+\frac{1}{2}})$, are constructed simultaneously, such that the following two properties hold. First, the vector fields Λ_{j+1} and

$$\tilde{\Delta}_j = \Lambda_j - \Lambda_{j+1} = \Phi_{j*}\Delta_j$$

are integrable. Second, the vector field $\check{X}_j$ defined on $\mathscr{D}_{j+1}$ that satisfies

$$\Psi_{j*}\check{X}_j = X_j + \Delta_j$$

has the property that its normal linear part $\mathsf{N}\check{X}_j$ is much closer to L_Ω than $\mathsf{N}X_j$, in a sense that will be made precise below. Note that, unlike the vector field $\tilde{\Delta}_j$, the vector field Δ_j need not and in general will not be integrable.

The coordinate transformation Φ_{j+1} is then obtained by setting

$$\Phi_{j+1} = \Phi_j \circ \Psi_j.$$

With the knowledge of Φ_{j+1}, the domain $\tilde{\mathscr{D}}_{j+1}$ is determined by (27), and the vector field X_{j+1} is determined by setting

$$(\Phi_{j+1})_* X_{j+1} = \tilde{X}_{j+1} - \Lambda_{j+1}.$$

Finally, we show that the limits

$$\tilde{X}_j - \Lambda_j \to \tilde{X} - \Lambda, \qquad X_j \to X, \qquad \Phi_j \to \Phi$$

exist as $j \to \infty$, that $\mathsf{N}X = L_\Omega$, and that

$$\Phi_* X = \tilde{X} - \Lambda.$$

Remark 3. Necessary for these constructions is that for all j:

$$\Phi_j(\mathscr{D}_j) \subset \tilde{\mathscr{D}}_j, \qquad \tilde{\mathscr{D}}_{j+1} \subset \Phi_j\left(\mathscr{D}_{j+\frac{1}{2}}\right). \tag{28}$$

The first inclusion ensures that the vector field X_j is defined on $\mathscr{D}_j$, and the second ensures that Λ_{j+1} is defined on all of $\tilde{\mathscr{D}}_{j+1}$.

4.3. **Approximation.** In order to construct analytic approximations $\tilde{X}_j$ of $\tilde{X}$ on the complex domains $\tilde{\mathscr{D}}_j$, a modified version of Zehnder's approximation technique (see [39]) is used, which gives explicit information on the growth of constants that depend on the degree of differentiability.

4.3.1. *Finite differentiability.* We need the following sharpened version of Zehnder's approximation lemma. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called periodic with periods T_i, $i = 1, ..., n$, if $f(x + T_i) = f(x)$ for all x and all i.

Lemma 4.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be r–times continuously differentiable, and let $\{\rho_j\}_{j=0}^\infty$ be a monotonically decreasing sequence of positive real numbers. For every $\eta \in (0, 1)$, and for every $j > 0$, there exists an entire holomorphic function $f_j : \mathbb{C}^n \to \mathbb{C}$, taking real values on real vectors, such that*

$$\|f_j - f\|_{\mathcal{C}^s} \to 0 \quad \text{as} \quad j \to \infty, \quad \text{for all } 0 \le s < r,$$

and

$$|f_j - f_{j-1}|_{\rho_j} \le c_1^{s+1} (s!)^\eta \rho_{j-1}^s \|f\|_{\mathcal{C}^s}, \quad \text{for every } 1 \le s \le r;$$

here $c_1 = 2\,\mathrm{e}^2 \left(\frac{2^6}{\eta}\right)^{2n}$.

If f is periodic in its argument, then every f_j can be chosen to be periodic with the same periods.

The proof of lemma 4.1 follows [39] closely; the main difference is that $\mathcal{C}^\infty$ bump functions are replaced by Gevrey regular bump functions.

The construction of these bump functions is the content of the next lemma. Then in lemma 4.1 the approximating functions are constructed by convolving f with the inverse Fourier transform φ of Gevrey bump functions $\hat{\varphi}$. Estimates on the derivatives of the smoothed functions are obtained in terms of the derivatives of $\hat{\varphi}$. Finally, the smoothing is applied repeatedly in different directions.

Lemma 4.2. *Let $0 < \eta < 1$. There exists an even, infinitely differentiable function $\psi : \mathbb{R} \to \mathbb{R}$, vanishing on the complement of the open interval $(-2, 2)$, taking the value 1 on the closed interval $[-1, 1]$, whose derivatives can be bounded as*

$$|\psi^{(s)}(x)| \leq \left(\frac{4}{\eta}\right)^s (s!)^{1+\eta}, \tag{29}$$

for all $x \in \mathbb{R}$, and all $s \geq 0$.

Proof. The function ψ is constructed by repeatedly convolving multiples of indicator functions (see e.g. [22]).

Introduce $a_k = c\,(k+1)^{-1-\eta}$ and choose c such that $\sum_{k=0}^{\infty} a_k = 1$. Since

$$\int_0^\infty \frac{1}{(x+1)^{1+\eta}}\,\mathrm{d}x \leq \sum_{k=0}^{\infty} \frac{1}{(k+1)^{1+\eta}} \leq 1 + \int_1^\infty \frac{1}{x^{1+\eta}}\,\mathrm{d}x,$$

it follows that $\frac{\eta}{1+\eta} \leq c \leq \eta$.

Introduce for $a > 0$ the function $H_a : \mathbb{R} \to \mathbb{R}$ by

$$H_a(x) = \begin{cases} a^{-1} & \text{for } x \in (0, a), \\ 0 & \text{otherwise.} \end{cases}$$

The convolution $u * v$ of two integrable functions $u, v : \mathbb{R} \to \mathbb{R}$ is given by

$$u * v = \int_{\mathbb{R}} u(x-y)v(y)\,\mathrm{d}y;$$

we have that $\int_{\mathbb{R}} u * v\,\mathrm{d}x = \int_{\mathbb{R}} u\,\mathrm{d}x \cdot \int_{\mathbb{R}} v\,\mathrm{d}x$. Using the sequence a_k, define a sequence of functions

$$u_k = H_{a_0} * H_{a_1} * \cdots * H_{a_k},$$

and note that $\int_{\mathbb{R}} u_k\,\mathrm{d}x = 1$ since $\int_{\mathbb{R}} H_a\,\mathrm{d}x = 1$. It follows from theorem 1.3.5 of [22] and the fact that $\sum a_k = 1$ that the sequence $\{u_k\}$ converges uniformly to a smooth function $u : \mathbb{R} \to \mathbb{R}$ with support in $[0, 1]$, which is such that $\int_{\mathbb{R}} u\,\mathrm{d}x = 1$ and

$$|u^{(s)}(x)| \leq \frac{(s+1)^{1+\eta}}{c}\left(\frac{2}{c}\right)^s (s!)^{1+\eta},$$

for all $s \in \mathbb{N}$. Note that $u \in \mathcal{G}^{1+\eta}$.

The function $v(t) = u(-x-1) - u(x-1)$ has support $[-2, -1] \cup [1, 2]$, it is odd, and $\int_{\mathbb{R}} v\,\mathrm{d}x = 0$. Hence, its primitive

$$\psi(x) = \int_{-\infty}^{x} v(t)\,\mathrm{d}t$$

is even, vanishes for all x in the complement of $[-2, 2]$, and satisfies $\psi(x) = 1$ for $|x| \leq 1$. Moreover, for $s \geq 1$,

$$\left|\psi^{(s)}(x)\right| \leq \frac{1}{2}\left(\frac{2}{c}\right)^s (s!)^{1+\eta},$$

and $\psi \in \mathcal{G}^{1+\eta}$. Using $c \geq \frac{\eta}{1+\eta} \geq \eta/2$ for $0 < \eta < 1$ yields the lemma. $\square$

We can now prove lemma 4.1. The proof consists of three parts: first we define holomorphic approximations f_j of f; then we show that these converge to f as $j \to \infty$, and finally we demonstrate the bound on the difference $|f_j - f_{j-1}|$.

Proof. Let $\hat\varphi$ be equal to the function ψ given by lemma 4.2, and let φ be its inverse Fourier transform, given as

$$\varphi(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{i\xi x}\hat\varphi(\xi)\,\mathrm{d}\xi.$$

The function $\hat\varphi$ is a Schwartz function, that is, $|x|^k\hat\varphi^{(s)}(x)$ is bounded for every $k, s > 0$; as the Fourier transformation interchanges differentiation and multiplication with a mononomial, the transformed function φ is a Schwartz function as well, and hence φ and all its derivatives are integrable. Moreover, since $\hat\varphi$ is even, the function φ maps $\mathbb{R}$ onto itself, it satisfies $\int_{\mathbb{R}}\varphi\,\mathrm{d}x = \hat\varphi(0) = 1$, and as $\hat\varphi$ has compact support, the function φ can be continued analytically to an entire function φ on $\mathbb{C}$.

For $t > 0$, introduce $\varphi_t(x) = t\varphi(tx)$; note that for every $t > 0$ the function φ_t has the same properties as those stated for φ in the previous paragraph. For every bounded continuous real–valued function f on $\mathbb{R}$, the analytic smoothing $S_t f$ of f is defined by

$$S_t f(z) = \int_{\mathbb{R}} \varphi_t(z-y)f(y)\,\mathrm{d}y. \tag{30}$$

The analytic smoothing of f is an entire holomorphic function on $\mathbb{C}$, taking real values on real arguments. It is easy to verify that if f is periodic, then so is $S_t f$, and for functions f with bounded derivatives, smoothing commutes with differentiation: for $s \in \mathbb{N}$ with $s < r$ we have $S_t f^{(s)} = (S_t f)^{(s)}$. The holomorphic approximations f_j are defined as

$$f_j = S_{\rho_j^{-1}} f,$$

where $\{\rho_j\}$ is the given monotonic sequence.

Let $s \in [0, r)$, and introduce $g = f^{[s]}$, where $[s]$ is the largest integer smaller than or equal to s. We wish to show convergence of $S_t g$ to g as $t \to \infty$ in the $\mathcal{C}^\alpha$-norm, where $0 < \alpha = s - [s] < r - [s] = \beta$. For this, note that $g \in \mathcal{C}^\beta(\mathbb{R})$. Fix $\delta > 0$ arbitrarily.

For $h > \delta$, we have that

$$\begin{aligned}
&\frac{\big|(g - S_t g)(x+h) - (g - S_t g)(x)\big|}{h^\alpha} \\
&= h^{-\alpha}\left|\int \varphi_t(-y)\Big(g(x+h) - g(y+x+h)\Big)\,\mathrm{d}y - \int \varphi_t(-y)\Big(g(x) - g(y+x)\Big)\,\mathrm{d}y\right| \\
&\le h^{-\alpha}\int_{|y|\le\delta} \varphi_t(y)\left(\Big|g(x+h+y) - g(x+h)\Big| + \Big|g(x+y) - g(x)\Big|\right)\,\mathrm{d}y \\
&\qquad + h^{-\alpha}\int_{|y|>\delta} \varphi_t(y)\Big(\big|g(x+h)\big| + \big|g(x+y+h)\big| + \big|g(x)\big| + \big|g(x+y)\big|\Big)\,\mathrm{d}y \\
&\le 2\|g\|_{\mathcal{C}^\beta}\frac{\delta^\beta}{h^\alpha} + 4\|g\|_{\mathcal{C}^\beta}h^{-\alpha}\int_{|y|>\delta}\varphi_t(y)\,\mathrm{d}y \\
&\le 6\|f\|_{\mathcal{C}^r}\delta^{r-s}.
\end{aligned}$$

For the first inequality, we used the fact that φ_t is even. The last inequality follows by choosing t so large that the integral on the one but last line is made smaller than δ^β.

For $0 < h \le \delta$, the following straightforward estimates hold:

$$\begin{aligned}
&\frac{\left|\big(g - S_t g\big)(x+h) - \big(g - S_t g\big)(x)\right|}{h^\alpha} \\
&\le \left|\frac{g(x+h) - g(x)}{h^\alpha}\right| + \int \varphi_t(-y) \left|\frac{g(y+x+h) - g(y+x)}{h^\alpha}\right| \mathrm{d}y \\
&\le 2\|g\|_{\mathcal{C}^\beta} h^{\beta-\alpha} \le 2\|f\|_{\mathcal{C}^r} \delta^{r-s}.
\end{aligned}$$

As $\delta > 0$ was arbitrary, $\|f - S_t f\|_{\mathcal{C}^s} \to 0$ as $t \to \infty$. This shows the first clause of lemma 4.1.

To show the second clause, introduce functions χ_s and ψ_s by

$$\chi_s(y) = \frac{1}{(s-1)!} \int_{\mathbb{R}} |\varphi(x - \mathrm{i}y)||x|^s \,\mathrm{d}x, \tag{31}$$

with the convention $(-1)! = 0! = 1$, and $\psi_s(\rho) = 2\sup_{|y|<\rho} \chi_s(y)$.

Let $\rho > 0$, and let $f \in \mathcal{C}^r(\mathbb{R})$. The following two estimates are taken from [39]. It is shown there that

$$\|S_t f - f\|_{\mathcal{C}^0} \le t^{-s} \chi_s(0) \|f\|_{\mathcal{C}^s}, \tag{32}$$

$$|S_{\rho^{-1}} f - S_t f|_\rho \le t^{-s} \psi_s(1) \|f\|_{\mathcal{C}^s}, \tag{33}$$

for all $0 < t \le \rho^{-1}$.

We need an explicit bound of $\chi_s(y)$ for all $|y| < 1$. Using (31) and the fact that $\hat{\varphi}$ is the Fourier transform of φ yields:

$$\begin{aligned}
\chi_s(y) &= \frac{1}{(s-1)!} \int_{\mathbb{R}} \left| \frac{1}{2\pi} \int_{\mathbb{R}} x^s \hat{\varphi}(\xi)\, \mathrm{e}^{\mathrm{i}\xi(x - \mathrm{i}y)} \,\mathrm{d}\xi \right| \mathrm{d}x \\
&\le \frac{\mathrm{e}^{2y}}{(s-1)!} \int_{\mathbb{R}} \left| \int_{\mathbb{R}} x^s \hat{\varphi}(\xi)\, \mathrm{e}^{\mathrm{i}\xi x} \,\mathrm{d}\xi \right| \mathrm{d}x.
\end{aligned}$$

The inequality follows since the support of $\hat{\varphi}$ is contained in $[-2, 2]$. By splitting the domain of integration over x, noting that the integrand is even in x, and repeated partial integration over ξ, the following estimate is obtained:

$$\begin{aligned}
\chi_s(y) \le \frac{2e^{2y}}{(s-1)!} &\int_0^1 \left| \int_{\mathbb{R}} \hat{\varphi}^{(s)}(\xi)\, \mathrm{e}^{-\mathrm{i}\xi x} \,\mathrm{d}\xi \right| \mathrm{d}x \\
&+ \frac{2e^{2y}}{(s-1)!} \int_1^\infty x^{-2} \left| \int_{\mathbb{R}} \hat{\varphi}^{(s+2)}(\xi)\, \mathrm{e}^{-\mathrm{i}\xi x} \,\mathrm{d}\xi \right| \mathrm{d}x.
\end{aligned}$$

Restricted to the support of $\hat{\varphi}$, the integrands are estimated using (29) and $0 \le \eta \le 1$, which yields

$$\psi_s(1) = 2 \sup_{|y|<1} \chi_s(y) \le 256\, \mathrm{e}^2 \left(1 + \frac{s}{2}\right)^4 \left(\frac{4}{\eta}\right)^{s+2} (s!)^\eta. \tag{34}$$

Let ρ_j be as in the statement of the lemma, and set $f_j = S_{\rho_j^{-1}} f$. Combining the estimate (34) with (33) yields

$$|f_j - f_{j-1}|_{\rho_j} \le \frac{4096\, \mathrm{e}^2}{\eta^2} \left(\frac{4\, \mathrm{e}^2}{\eta}\right)^s \rho_{j-1}^s (s!)^\eta \|f\|_{\mathcal{C}^s}.$$

It follows immediately from (30) and the definition of $\chi_s(y)$ that

$$\sup_{\mathbb{R}} |S_t f^{(s)}(x)| \le \chi_0(0) \sup_{\mathbb{R}} |f^{(s)}(x)|.$$

Consequently

$$\|S_t f\|_{\mathcal{C}^s} \le \chi_0(0)\|f\|_{\mathcal{C}^s}. \tag{35}$$

Consider now a $\mathcal{C}^r$ function $f : \mathbb{R}^n \to \mathbb{R}$. To ease notation, let S_j^i denote the smoothing operator $S_{\rho_j^{-1}}$ in the direction of x_i. Introduce $f_j = S_j^1 \cdots S_j^n f$, and estimate

$$\begin{aligned}
|f_j - f_{j-1}|_{\rho_j} &\le |S_j^1 S_j^2 \cdots S_j^n f - S_{j-1}^1 S_j^2 \cdots S_j^n f|_{\rho_j} \\
&\qquad + \cdots + |S_{j-1}^1 \cdots S_{j-1}^{n-1} S_j^n f - S_{j-1}^1 \cdots S_{j-1}^{n-1} S_{j-1}^n f|_{\rho_j} \\
&\le \rho_{j-1}^s \psi_s(1)\|S_j^2 \cdots S_j^n f\|_{\mathcal{C}^s} + \cdots + \rho_{j-1}^s \psi_s(1)\|S_{j-1}^1 \cdots S_{j-1}^{n-1} f\|_{\mathcal{C}^s} \\
&\le n(\chi_0(0))^{n-1}\psi_s(1)\rho_{j-1}^s \|f\|_{\mathcal{C}^s} \\
&\le n(\chi_0(0))^{n-1}\psi_s(1)\rho_{j-1}^s \|f\|_{\mathcal{C}^s}.
\end{aligned}$$

The second inequality follows from (33) and the fact that the smoothing operators in the different directions commute, and the final inequality follows from equation (35). As $\chi_0(0) \le 2^{11}/\eta^2$, we obtain that

$$|f_j - f_{j-1}|_{\rho_j} \le 2\,\mathrm{e}^2 n \left(\frac{2^{11}}{\eta^2}\right)^n \left(\frac{4\,\mathrm{e}^2}{\eta}\right)^2 (s!)^\eta \rho_{j-1}^s.$$

This implies the lemma in the general case. □

4.3.2. *Legendre transformation.* Using the approximation result obtained in 4.3.1, we derive a variant that yields holomorphic approximations to functions in a Carleman class $\mathcal{C}^M$. As a preparation, some concepts from the theory of convexity are recalled.

Let $f : [0, \infty) \to \mathbb{R}$ be an increasing convex function. Define the Legendre transform $\mathcal{L}f$ of f as follows: for every $p > 0$, the value $g(p) = \mathcal{L}f(p)$ is the smallest q such that

$$f(x) \ge px - q \text{ for all } x > 0.$$

and $f(\bar{x}) = p\bar{x} - g(p)$ for some $\bar{x} > 0$. If equality holds and f is differentiable at $\bar{x}$, then $p = f_*'(\bar{x})$.

The function g is also convex. Moreover, if $\lim_{x\to\infty} f(x)/x = \infty$, then the gradient $f'(x) = p$ of f tends to infinity as $x \to \infty$, $g(p)$ is defined for all $p > 0$, and

$$\lim_{p\to\infty} g(p)/p = \infty$$

as well.

As an example we calculate the Legendre transformation of $f(x) = a\,\mathrm{e}^{bx} - cx$, which will be needed later. Since f is differentiable,

$$p = f'(x) = ab\,\mathrm{e}^{bx} - c.$$

Solving for x yields that $x = (1/b)\log((p + c)/(ab))$. We find g by substitution:

$$g(p) = xp - f(x) = \frac{p + c}{b}\left(\log\frac{p + c}{ab} - 1\right). \tag{36}$$

In general, if f is convex, left- and right-hand limits of the derivative f' exist at every point $x > 0$. The interval $\partial f(x_0) = [\lim_{x\uparrow x_0} f'(x), \lim_{x\downarrow x_0} f'(x)]$ is called the subgradient of f. If the graph of f has a corner, that is, if x_0 is such that the subgradient $\partial f(x_0)$ has nonempty interior, then for $p \in \partial f(x_0)$:

$$g(p) = x_0 p - f(x_0)$$

and $g'(p) = x_0$.

Let $\{f_n\}_{n=0}^\infty$ be an increasing sequence of real numbers. The largest convex minorant f_* of $\{f_n\}$ is the function $f_* : [0, \infty) \to \mathbb{R}$ such that

$$f_*(x) = \inf_{a,b\in\mathbb{N}} \left\{ \frac{b-x}{b-a} f_a + \frac{x-a}{b-a} f_b, \text{ if } a < x < b \right\};$$

put differently, f_* is that function for which the epigraph $\{(x, y) \in \mathbb{R}^2 \mid x \geq 0, y \geq f_*(x)\}$ equals the convex hull of set formed of the points (i, f_i) and the positive vertical axis.

4.3.3. *Ultradifferentiability.* Lemma 4.1 will now be applied to elements of the Carleman classes $\mathcal{C}^M$. Let $M = \{M_s\}$ be the increasing sequence of positive real numbers M_s defining the class. The space of $\mathcal{C}_h^M$-smooth vertical vector fields on $\mathscr{M} \times \mathscr{P}$ will be denoted by $\mathcal{X}_h^M$.

From the definition of the Carleman classes (9), it follows that if $f \in \mathcal{C}_h^M$, then

$$\|f\|_s \leq Ch^{-s} M_s \qquad \text{for every } s \in \mathbb{N}.$$

We obtain from lemma 4.1 that there exists a sequence of entire holomorphic functions f_j, converging to f in every $\mathcal{C}^s$-norm, and such that

$$|f_j - f_{j-1}|_{\rho_j} \leq C_0^{s+1} (s!)^\eta \rho_{j-1}^s M_s$$

for all s, where $C_0 = \max\{c_1/h, C\}$. Let $0 < \eta < 1$ be a given constant, and let c_1 be as in lemma 4.1. Let $\lambda : [0, \infty) \to \mathbb{R}$ be any strictly increasing convex function, such that for $s \in \mathbb{N}$

$$\lambda(s) \geq \lambda_s = \log C_0 + s \log C_0 + \eta \log s! + \log M_s. \tag{37}$$

Note that we could take for λ the largest convex minorant λ_* of the sequence $\{\lambda_s\}$, since for every other function λ satisfying the conditions we have $\lambda(s) \geq \lambda_*(s)$. It is however convenient, when dealing with the Gevrey class, to be able to work with differentiable functions λ.

Recall that the domains $\mathscr{D}_j$ are defined in terms of the decreasing sequence $\{r_j\}$ in (26).

Lemma 4.3. *The sequence $\{b_j\}_{j=1}^\infty$, given by*

$$b_j = \exp(-\mathcal{L}\lambda(\log r_{j-1}^{-1})),$$

satisfies for all $s > 0$

$$\lim_{j\to\infty} b_j / r_j^s = 0,$$

and for any vector field $\tilde{X} \in \mathcal{X}_h^M$, there is a sequence of approximating holomorphic vector fields $\tilde{X}_j$ such that

$$|\tilde{X}_j - \tilde{X}_{j-1}|_{\tilde{\mathcal{D}}_j} \leq b_j \|\tilde{X}\|_{\mathcal{X}_h^\mu}.$$

Proof. Lemma 4.1 gives for $\tilde{X} \in \mathcal{X}_h^M$ a sequence of entire holomorphic vector fields $\{\tilde{X}_j\}$ which converge uniformly to $\tilde{X}$ in $\mathcal{X}^s$ for every $s \in \mathbb{N}$. Moreover, there is the estimate

$$|\tilde{X}_j - \tilde{X}_{j-1}|_{\tilde{\mathcal{D}}_j} \leq C_0^{s+1} (s!)^\eta r_{j-1}^s M_s \|\tilde{X}\|_{\mathcal{X}_h^M} \leq r_{j-1}^s \, \mathrm{e}^{\lambda(s)} \|\tilde{X}\|_{\mathcal{X}_h^M},$$

which holds for every $s \in \mathbb{N}$. The left hand side of the inequality does not depend on s.

Let s_j be the smallest value of s such that the right hand side of the inequality is minimal, that is, such that for all $s \geq 0$:

$$r_{j-1}^{s_j} \, \mathrm{e}^{\lambda(s_j)} \leq r_{j-1}^{s} \, \mathrm{e}^{\lambda(s)}.$$

Set

$$b_j = r_{j-1}^{s} \, \mathrm{e}^{\lambda(s_j)} \leq r_{j-1}^{s} \, \mathrm{e}^{\lambda(s)}.$$

Taking logarithms of this inequality yields that

$$\lambda(s) \geq \log b_j + s \log r_j^{-1};$$

moreover, equality holds if $s = s_j$. This is exactly the formulation of the Legendre transform, and we find that

$$\log b_j = -\mathcal{L}\lambda(\log r_{j-1}^{-1}).$$

Since $\lim_{s\to\infty} \lambda(s)/s = \infty$, we have that $\lim_{p\to\infty} \mathcal{L}\lambda(p)/p = \infty$. Making use of the fact that $\{r_j\}$ is a decreasing geometric sequence, we find for fixed $s \in \mathbb{N}$ that

$$\frac{b_j}{r_j^s} = \exp\left(\log r_j^{-1}\left(s - \frac{\log r_{j-1}^{-1}}{\log r_j^{-1}} \frac{\mathcal{L}\lambda(\log r_{j-1}^{-1})}{\log r_{j-1}^{-1}}\right)\right) \to 0 \quad \text{as } j \to \infty.$$

□

For the Gevrey class $\mathcal{G}_h^\mu$, the constants M_s equal $(s!)^\mu$ with $\mu > 1$, and f can be taken equal to

$$\lambda(s) = (\mu + \eta) s \log s + (s+1) \log c_1,$$

for some $\eta > 0$. We find

$$\log b_j = -\mathcal{L}\lambda(\log r_{j-1}^{-1}) = -C r_{j-1}^{-\frac{1}{\mu+\eta}} - \log c_1,$$

where $C = (\mu + \eta)\, \mathrm{e}^{-1-\log c_1/(\mu+\eta)}$. Consequently

$$|\tilde{X}_j - \tilde{X}_{j-1}|_{\tilde{\mathcal{D}}_j} \leq c_1 \exp\left(-C r_{j-1}^{-\frac{1}{\mu+\eta}}\right) \|\tilde{X}\|_{\mathcal{X}_h^\mu}. \tag{38}$$

4.3.4. *Application.* Recall from subsection 4.1, that the vector field $\tilde{X}$ can be written in the form

$$\tilde{X} = Z + P = L_\Omega + \tilde{Q} + P,$$

where $L_\Omega = \omega(p)\frac{\partial}{\partial x} + A(p)y\frac{\partial}{\partial y}$, and where $\tilde{Q} = q_1(y,p)\frac{\partial}{\partial x} + q_2(y,p)\frac{\partial}{\partial y}$ is integrable and such that $\mathsf{N}\tilde{Q} = 0$. The map L_Ω is real analytic; the (vertical) vector fields $\tilde{Q}$ and P are in the smoothness class $\mathcal{X}(\mathscr{M} \times \mathscr{P})$.

In the case that $\tilde{X}$ is itself real analytic, take $\tilde{X}_j = \tilde{X}$ for all j.

For the other cases, lemmas 4.1 and 4.3 yield a sequence $\{b_j\}$, which is determined only by the smoothness class, and holomorphic vector fields $\tilde{Q}_j$ and P_j of $\tilde{Q}$ and P respectively, defined on $\tilde{\mathcal{D}}_j$, that satisfy

$$|\tilde{Q}_j - \tilde{Q}_{j-1}|_{\tilde{\mathcal{D}}_j} \leq b_j \|\tilde{Q}\|_{\mathcal{C}^s}, \quad \text{and} \quad |P_j - P_{j-1}|_{\tilde{\mathcal{D}}_j} \leq b_j \|P\|_{\mathcal{C}^s}. \tag{39}$$

Here $b_j = c_s r_j^s$ in the case that $\mathcal{B} = \mathcal{C}^s$, and b_j is given by lemma 4.3 if $\mathcal{B} = \mathcal{C}_s^M$. Note that in general the normal linear part $\mathsf{N}\tilde{Q}_j$ of $\tilde{Q}_j$ will not vanish identically.

In both cases, define

$$\tilde{X}_j = L_\Omega + \tilde{Q}_j + P_j,$$

and note that the vector fields $\tilde{X}_j$ are holomorphic and tend to $\tilde{X}$ as $j \to \infty$. This concludes the first stage of the proof.

4.4. **The induction step.** This subsection treats the second stage of the proof, the inductive construction of the embedding Φ_{j+1} and the vector fields X_{j+1}, Δ_j and $\tilde{\Delta}_j$. At the beginning of the construction, an embedding $\Phi_j : \mathscr{D}_j \to \tilde{\mathscr{D}}_j$ and a vector field X_j on $\mathscr{D}_j$ are given.

As sketched in subsection 4.2, the aim of the induction step is to construct an embedding Φ_{j+1} and an integrable vector field $\tilde{\Delta}_j$, such that the normal linear part of the vector field X_{j+1} that satisfies

$$(\Phi_{j+1})_* X_{j+1} = \tilde{X}_{j+1} - \Lambda_j + \tilde{\Delta}_j$$

is "much" closer to $L = L_\Omega = \omega \frac{\partial}{\partial x} + Ay\frac{\partial}{\partial y}$ than $\mathsf{N}X_j$. If X_j is written as

$$X_j = L + R_j + Q_j, \tag{40}$$

where Q_j is such that $\mathsf{N}X_j = L + R_j$ and $\mathsf{N}Q_j = 0$, the 'distance' between $\mathsf{N}X_j$ and L can be expressed by the size of R_j. We shall demonstrate that $|R_j|_j \to 0$ as $j \to \infty$; moreover, the speed of this convergence is linked to the smoothness of the limiting embedding $\Phi_\infty = \lim_{j\to\infty} \Phi_j$.

4.4.1. *Induction assumptions.* We begin by stating the induction hypothesis precisely. It is assumed that embeddings Ψ_1, ..., Ψ_{j-1}, Φ_1, ..., Φ_j and vector fields X_1, ..., X_j, Λ_1, ..., Λ_j are already constructed as indicated in subsection 4.2. All embeddings and all vector fields are complex extensions of real analytic ones, taking real values when restricted to real vectors.

To formulate the assumptions, introduce maps φ_i and ψ_i by setting $\Phi_i = \mathrm{id}_{\mathscr{D}_i} + \varphi_i$ and $\Phi_i^{-1} = \mathrm{id}_{\Phi(\mathscr{D}_i)} + \psi_i$, and define maps $(\varphi_i)_p$ and $(\psi_i)_p$ taking values in $\mathbb{T}^m \times \mathbb{R}^n$ by setting $\varphi_i(x, y, p) = ((\varphi_i)_p(x, y), 0)$ etc.

Hypothesis. *There is a constant $c \in (0, 1)$, not depending on j, such that*

$$|\Psi_i - \mathrm{id}_{\mathscr{D}_{i+1}}|_{i+1} < c\, r_{i+1}, \qquad \textit{for} \quad 1 \le i \le j-1, \tag{41}$$

and such that

$$|\varphi_i|_i, |D\varphi_i|_i, |D\psi_i|_{\Phi_i(\mathscr{D}_i)} < c - \frac{r_i^2}{r_i - r_{i+1}}, \tag{42}$$

for all $1 \le i \le j$. Moreover, there is a constant $C > 0$, also not depending on j, such that for R_j and Q_j as in (40),

$$|R_j|_j < \frac{1}{C} \frac{r_j^{(n^2+n)\kappa+3}}{d_j^{n^2+n}} \quad \textit{and} \quad |Q_j|_j \le (2 - 2^{-j+1})|Q_1|_1. \tag{43}$$

Finally, the vector fields $\tilde{\Delta}_i$ are integrable for all $1 \le i \le j-1$.

Note that the hypothesis holds for the case $j = 1$, with $X_1 = \tilde{X}_1$ and $\Phi_1 = \mathrm{id}_{\mathscr{D}_1}$.

4.4.2. *'+' and '·'–notation.* In order not to overburden the notation, so–called '+'–notation will be used. All indices 'j' are dropped, and indices '$j + 1$' are replaced by '+'. In this notation, the vector field $X_j + \Delta_j$ defined on $\mathscr{D}_j$ is written as $X + \Delta$, defined on $\mathcal{D}$.

In the estimates below, also the so–called '·'–notation will be used. Whenever $s <\cdot\ t$ is written, it is taken to signify $s < Mt$, where the constant M does not depend on j.

4.4.3. *Inclusion of domains.* According to the sketch of the proof given in 4.2, see in particular Remark 3, we should have that $\tilde{\mathscr{D}}_+ \subset \Phi(\mathscr{D}_{\frac{1}{2}})$ and $\Phi(\mathscr{D}) \subset \tilde{\mathscr{D}}$. We shall require a little bit more.

Recall that $\mathscr{V}$ is a bounded real neighbourhood of $\mathscr{T} = \mathbb{T}^m \times \{0\}$, and that $\mathscr{U}$ equals the complex neighbourhood $\mathscr{V} + h$ in the real analytic case, and $\mathscr{V}$ otherwise.

Lemma 4.4. *Assume that the inequalities*

$$r + 2cr < \tilde{r}, \quad \tilde{r}_+ < (1-c)(\tfrac{1}{2}r + \tfrac{1}{2}r_+) \tag{44}$$

are satisfied, together with the induction assumptions. Then the inclusions

$$\Phi_p(\mathscr{U} + r) \subset \Phi_p(\mathscr{U}) + \tilde{r} - cr \quad \textit{and} \quad \Phi_{p+}(\mathscr{U}) + \tilde{r}_+ \subset \Phi_p(\mathscr{U} + r_{\frac{1}{2}})$$

hold true. Also, if

$$\rho + 2c\rho < \tilde{\rho}, \quad \tilde{\rho}_+ + 2c\tilde{\rho}_+ < \rho \quad \textit{and} \quad \rho < r, \quad \tilde{\rho} < \tilde{r}, \tag{45}$$

then

$$\mathscr{P}(\tilde{\rho}_+) \subset \mathscr{P}(\rho) \subset \mathscr{P}(\tilde{\rho} - c\rho).$$

Proof. The second clause is immediate. The first clause is a direct consequence of the induction hypothesis; this can be seen as follows. For the first inclusion, take $z = z_0 + z_1 \in \mathscr{U} + r$ such that $z_0 \in \mathscr{U}$ and $|z_1| < r$. Then by the mean value theorem, there is $\vartheta \in (0,1)$ such that for $z_\vartheta = z_0 + \vartheta z_1$:

$$\Phi_p(z) = \Phi_p(z_0) + D\Phi_p(z_\vartheta)z_1 = \Phi_p(z_0) + z_1 + D\varphi_p(z_\vartheta)z_1. \tag{46}$$

Since

$$|z_1 + \varphi_p(z_\vartheta)z_1| < r + cr,$$

the condition $r + 2cr < \tilde{r}$ implies that $\Phi_p(z) \in \Phi_p(\mathscr{U}) + \tilde{r} - cr$.

To see the second inclusion, take $z = z_0 + z_1 \in \mathscr{U} + r$ such that $z_0 \in \mathscr{U}$, z is on the boundary of $\mathscr{U} + r_{\frac{1}{2}}$ and that the norm $|z_1|$ of z_1 is minimal, and therefore equal to $|z_1| = r_{\frac{1}{2}}$. With the same notation as before, again (46) holds. We conclude that the distance from $\Phi_p(z)$ to $\Phi_p(\mathscr{U})$ is bounded from below by

$$|z_1| - \max|D\varphi_p||z_1| > (1-c)|z_1| = (1-c)(\tfrac{1}{2}r + \tfrac{1}{2}r_+).$$

Consequently any point in the set $\Phi_p(\mathscr{U}) + \tilde{r}_+$ is necessarily in the interior of the set $\Phi_p(\mathscr{U} + r_{\frac{1}{2}})$. □

We shall assume that $\{r_j\}$, $\{\tilde{r}_j\}$, $\{\rho_j\}$ and $\{\tilde{\rho}_j\}$ are decreasing geometric sequences; in particular, we set

$$\frac{\tilde{r}_+}{\tilde{r}} = \frac{r_+}{r} = a_1, \qquad \frac{\tilde{\rho}_+}{\tilde{\rho}} = \frac{\rho_+}{\rho} = a_2, \tag{47}$$

for some $0 < a_1, a_2 < 1$, and write $\tilde{r}_j = \tilde{r}_0 a_1^j$, $r_j = r_0 a_1^j$, *etc.* In terms of these constants, the inequalities (44) are equivalent to

$$1 + 2c < \frac{\tilde{r}_0}{r_0} < (1-c)\left(\frac{1}{2a_1} + 1\right). \tag{48}$$

Necessarily the constant c should be so small that

$$\frac{1+2c}{1-c} < \frac{1}{2a_1} + 1;$$

note that for any given a_1, such a $c > 0$ exists, since the left hand side of this inequality tends to 1 as $c \downarrow 0$.

4.4.4. *Form of the correction term.* The vector field $\tilde{\Delta}$ on $\tilde{\mathscr{D}}$ is taken to be of the form

$$\tilde{\Delta} = \delta \frac{\partial}{\partial x} + (b + By) \frac{\partial}{\partial y};$$

the 'modifying terms' $\delta(p) \in \mathbb{C}^m$, $b(p) \in \mathbb{C}^n$ and $B(p) \in \mathfrak{g}_{\mathbb{C}}$ taking real values on real vectors. Note that since Λ is integrable, the vector field $\Lambda_+ = \Lambda + \tilde{\Delta}$ will be integrable as well.

The vector field Δ on $\mathscr{D}$ is the image of $\tilde{\Delta}$ under the inverse of the already known map Φ; it can be written in the form

$$\Delta = \Phi_*^{-1} \left(\delta \frac{\partial}{\partial x} + (b + By) \frac{\partial}{\partial y} \right) = \delta \frac{\partial}{\partial x} + (b + By) \frac{\partial}{\partial y} + \Theta,$$

where $|\Theta| \leq |D\psi|_{\Phi(\mathscr{D})}(|\delta| + |b| + |B|)$.

4.4.5. *Form of the conjugacy.* The new conjugacy Φ_+ will be of the form $\Phi_+ = \Phi \circ \Psi$; given Ψ, introduce

$$\check{X} = \Psi_*^{-1}(X + \Delta).$$

The conjugacy Ψ is taken as the time-1 map e^{-Y} of a real analytic average–0 vector field $-Y \in \mathfrak{h}$, defined on $\mathscr{D}$ and written as

$$Y = u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} = u(x, p) \frac{\partial}{\partial x} + (v_0(x, p) + v_1(x, p) y) \frac{\partial}{\partial y}.$$

Requiring Y to be of average–0 (over $\mathbb{T}^m$) is equivalent to require the coefficient functions to satisfy $[u]_{\mathbb{T}^m} = 0$ and $[v]_{\mathbb{T}^m} = 0$, where $[f]_{\mathbb{T}^m} = \int_{\mathbb{T}^m} f(x)\,\mathrm{d}x$.

Recall that the Lie bracket of two vector fields $Z_1 = a_1 \frac{\partial}{\partial x} + b_1 \frac{\partial}{\partial y}$, $Z_2 = a_2 \frac{\partial}{\partial x} + b_2 \frac{\partial}{\partial y}$ is given as

$$\begin{aligned}[Z_1, Z_2] = &\left(a_1 \frac{\partial a_2}{\partial x} + b_1 \frac{\partial a_2}{\partial y} - a_2 \frac{\partial a_1}{\partial x} - b_2 \frac{\partial a_1}{\partial y} \right) \frac{\partial}{\partial x} \\ &+ \left(a_1 \frac{\partial b_2}{\partial x} + b_1 \frac{\partial b_2}{\partial y} - a_2 \frac{\partial b_1}{\partial x} - b_2 \frac{\partial b_1}{\partial y} \right) \frac{\partial}{\partial y}.\end{aligned}$$

We have $\Psi^{-1} = \exp(Y)$ and

$$\begin{aligned}\check{X} &= \Psi_*^{-1}(X + \Delta) \\ &= \exp(Y)_*(L + R + Q + \Delta) \\ &= L + R + Q + \Delta + [L, Y] + [R + \Delta, Y] + [Q, Y] + S \qquad (49)\end{aligned}$$

where

$$S = \int_0^1 (1 - s)[[X + \Delta, Y], Y]_{\exp(sY)}\,\mathrm{d}s.$$

The coefficient functions u, v_0 and v_1 of Y will be chosen as trigonometric polynomials in x.

For any vertical vector field Z on $\mathscr{D}$, introduce the Fourier tructation $T_d Z$ to order d. That is, if $Z = \sum_{k \in \mathbb{Z}^m} Z_k(y, p)\,\mathrm{e}^{\mathrm{i}\langle k, x\rangle}$, let

$$T_d Z = \sum_{|k| \leq d} Z_k(y, p)\,\mathrm{e}^{\mathrm{i}\langle k, x\rangle}.$$

The vector fields Δ and Y are determined by the requirement that they annihilate the contribution of the term $T_d R$ in $\check{Z}$; that is, they are taken to solve the *homological equation*

$$[L, Y] + T_d \mathsf{N}[Q, Y] + T_d(R + \Delta) = 0. \tag{50}$$

Note that this is an equation in $\mathfrak{h}$. Under the assumption that (50) holds, using (49) and writing $\check{X} = L + \check{R} + \check{Q}$ with $\mathsf{N}\check{X} = L + \check{R}$, it follows that

$$\begin{aligned}
\check{R} &= [R + \Delta, Y] + (R + \Delta + \mathsf{N}[Q, Y] - T_d(R + \Delta + \mathsf{N}[Q, Y])) + \mathsf{N}S, \\
\check{Q} &= Q + [Q, Y] - N[Q, Y] + S - NS.
\end{aligned}$$

In the next subsections, equation (50) is solved and estimates for $\check{R}$ and $\check{Q}$ are given.

4.4.6. *Determining the conjugacy.* The techniques of solving the homological equation (50) are mostly well–known and only brief indications are given. However, the determination of the modifying terms δ, b and B requires some care.

Set $T_d R = f\frac{\partial}{\partial x} + (g_0 + g_1 y)\frac{\partial}{\partial y}$, $Q = q_1 \frac{\partial}{\partial x} + q_2 \frac{\partial}{\partial y}$ and

$$\begin{aligned}
T_d \Delta &= \delta \frac{\partial}{\partial x} + (b + By)\frac{\partial}{\partial y} + T_d \Theta \\
&= \delta \frac{\partial}{\partial x} + (b + By)\frac{\partial}{\partial y} + \tilde{\delta}\frac{\partial}{\partial x} + (\tilde{b} + \tilde{B}y)\frac{\partial}{\partial y},
\end{aligned}$$

Here $f(x, p)$, $g_0(x, p)$ and $g_1(x, p)$ are trigonometric polynomials in x, taking real values on real vectors; the functions $\tilde{\delta}$, $\tilde{b}$ and $\tilde{B}$ are also trigonometric polynomials in x; moreover, they depend analytically on p as well as on (δ, b, B), and they satisfy estimates of the form

$$|\tilde{\delta}| \le |D\psi|_{\Phi(\mathscr{D})}|\delta|, \quad |\tilde{b}| \le |D\psi|_{\Phi(\mathscr{D})}|b|, \quad |\tilde{B}| \le |D\psi|_{\Phi(\mathscr{D})}|B|; \tag{51}$$

the analytic functions q_1 and q_2 satisfy $q_1 = \mathsf{O}(|y|)$ and $q_2 = \mathsf{O}(|y|^2)$.

Equation (50) can be split into three components:

$$\omega\frac{\partial u}{\partial x} + \delta + \tilde{\delta} + T_d\left(v_0 \frac{\partial q_1}{\partial y}\right) = -f, \tag{52}$$

$$\omega\frac{\partial v_0}{\partial x} - A v_0 + b + \tilde{b} = -g_0, \tag{53}$$

$$\omega\frac{\partial v_1}{\partial x} - \operatorname{ad}_A v_1 + B + \tilde{B} + T_d\left(q_1 \frac{\partial v_0}{\partial x} + v_0 \frac{\partial q_2}{\partial y}\right) = -g_1. \tag{54}$$

Here $\operatorname{ad}_A v_1 = [A, v_1] = A v_1 - v_1 A$. In the following, we set

$$\tilde{q}_1 = T_d(v_0 \frac{\partial q_1}{\partial y}) \text{ and } \tilde{q}_2 = T_d(q_1 \frac{\partial v_0}{\partial x} + v_0 \frac{\partial q_2}{\partial y}).$$

Equations (52)–(54) are solved in three steps. First v_0 and b will be determined from equation (53), as functions of (x, p, δ, B) and (p, δ, B) respectively. Then δ and B will be determined from equations (52) and (54), and finally u and v_1 are obtained from the same equations.

Equation (53) is equivalent to the following relations between the Fourier coefficients of v_0 and g_0:

$$\begin{aligned}
b + [\tilde{b}]_{\mathbb{T}^m} &= -g_{00}, \\
\mathrm{i}\langle k, \omega\rangle v_{0k} - A v_{0k} &= -g_{0k} - \tilde{b}_k, \qquad \text{for} \quad 0 < |k| \le d;
\end{aligned}$$

recall that $v_{00} = 0$ since Y is average–0. The first equation is solved by using the implicit function theorem together with the estimates (42) and (51), which yields $b = \hat{b}(p, \delta, B)$. Note that the estimate (59) below will imply that the second equation can be solved on $\mathscr{D}$, and that it yields an analytic solutions $\hat{v}_{0k}$:

$$\hat{v}_{0k}(p, \delta, B) = -\left(\mathrm{i}\langle k, \omega\rangle I - A\right)^{-1}\left(g_{0k}(p) + \tilde{b}_k(p, \delta, \hat{b}, B)\right), \tag{55}$$

for $0 < |k| \leq d$, and $v_{0k} = 0$ otherwise.

Averaging equations (52) and (54) leads to

$$\begin{aligned} \delta + [\tilde{\delta}]_{\mathbb{T}^m} + [\tilde{q}_1]_{\mathbb{T}^m} &= -f_0 \\ B + [\tilde{B}]_{\mathbb{T}^m} + [\tilde{q}_2]_{\mathbb{T}^m} &= -g_{10} \end{aligned}$$

where everywhere $\hat{b}(p, \delta, B)$ is substituted for b. Applying the implicit function theorem again yields solutions $\delta = \delta(p)$ and $B = B(p)$. Substituting these in $\hat{b}$ and $\hat{v}_{0k}$ yields $b(p)$ and $v_{0k}(p)$.

Finally equations (52) and (54) are solved for the case $0 < |k| \leq d$; this yields

$$u_k = -\frac{\tilde{\delta}_k + \tilde{q}_{1k} + f_k}{\mathrm{i}\langle k, \omega\rangle}, \tag{56}$$

$$v_{1k} = -\left(\mathrm{i}\langle k, \omega\rangle I - \mathrm{ad}_A\right)^{-1}\left(\tilde{B}_k + \tilde{q}_{2k} + g_{1k}\right). \tag{57}$$

As before, estimate (59) and (60) imply that these solutions are bounded analytic functions.

Note that the vector field $Y = u\frac{\partial}{\partial x} + (v_0 + v_1 y)\frac{\partial}{\partial y}$ is a linear combination of vector fields in $\mathfrak{h}$, and therefore $Y \in \mathfrak{h}$.

4.4.7. *Estimates.* The truncation level d is chosen as follows

$$d = \frac{1}{2}\left(\frac{\gamma}{2\gamma_0\rho}\right)^{\frac{1}{\kappa+1}}. \tag{58}$$

We take ρ_0 sufficiently small as to ensure that $d_1 \geq 2$. Let (ω, α_A) be the frequency vector of (ω, A). Since $p \in \mathcal{P}(\rho)$, the frequency vector can be written in the form $(\omega, \alpha_A) = (\omega_0, \alpha_0) + (\omega_1, \alpha_1)$ with (ω_0, α_0) normally Diophantine and $|\omega_1| + |\alpha_1| < \rho$. Hence, for $0 < |k| \leq d$ and $|\ell| \leq 2 \leq d$,

$$\begin{aligned} &|\mathrm{i}\langle k, \omega\rangle + \mathrm{i}\langle \ell, \alpha\rangle| \\ &\quad\geq \frac{\gamma}{\gamma_0}(|k| + |\ell|)^{-\kappa} - 2d\rho \\ &\quad\geq (|k| + |\ell|)^{-\kappa}\left(\frac{\gamma}{\gamma_0} - \rho(2d)^{\kappa+1}\right) \\ &\quad\geq \frac{\gamma}{2}(|k| + |\ell|)^{-\kappa}. \end{aligned} \tag{59}$$

Likewise, we obtain for $0 < |k| \leq d$ that

$$|i\langle k, \omega\rangle| \geq \frac{1}{2}|k|^{-\kappa}. \tag{60}$$

From estimates (59) and (60) it follows that on the open set $\mathcal{P}(\rho)$ the normal Diophantine conditions hold for those resonances whose order k satisfies $|k| \leq d$. As mentioned, this implies that all formal solutions given above are in fact well–defined analytic functions.

Recall that Cramer's rule allows us to express the inverse of a matrix A as

$$A^{-1} = (\det A)^{-1} A^*,$$

where A^* is the adjoint of A, that is, the matrix whose (i,j)'th element is the minor of the matrix obtained from A by removing the i'th row and the j'th column. We have to invert the linear maps $\mathrm{i}\langle k,\omega\rangle I - A$ and $\mathrm{i}\langle k,\omega\rangle I - \mathrm{ad}_A$. If λ_i, $i = 1, \cdots, n$ are the eigenvalues of A, then the eigenvalues of these maps are

$$\mathrm{i}\langle k,\omega\rangle I - \lambda_i \quad \text{and} \quad \mathrm{i}\langle k,\omega\rangle I - (\lambda_{i_1} - \lambda_{i_2}) \quad \text{respectively,}$$

for $i, i_1, i_2 \in \{1, \cdots, n\}$. The matrix elements of the adjoint to these maps contain terms with at most n factors $\langle k,\omega\rangle$ in the first case, and n^2 such factors in the second case.

Using Cramer's rule, and Rüssmann's technique to obtain optimal estimates (cf. [31, 32]), for b and v_0 the following inequalities are obtained:

$$|b| <\cdot\ |R|, \quad |v_0|_{\frac{1}{4}} <\cdot\ d^n \frac{|R|}{(r - r_{\frac{1}{4}})^{n\kappa}} <\cdot\ d^n \frac{|R|}{r^{n\kappa}};$$

Using these, a second application of Cramer's rule and Rüssmann's estimates yields for u, v_1, δ and B:

$$\begin{aligned}
|\delta|_{\frac{1}{2}} &<\cdot\ d^n \frac{|R|(1+|Q|)}{(r - r_{\frac{1}{2}})(r - r_{\frac{1}{4}})^{n\kappa}} && <\cdot\ d^n \frac{|R|}{r^{n\kappa+1}}, \\
|u|_{\frac{3}{4}} &<\cdot\ \frac{d^n |R|(1+|Q|)}{(r - r_{\frac{1}{2}})(r - r_{\frac{1}{4}})^{n\kappa}(r_{\frac{1}{2}} - r_{\frac{3}{4}})^{\kappa}} && <\cdot\ d^n \frac{|R|}{r^{(n+1)\kappa+1}}, \\
|B|_{\frac{1}{2}} &<\cdot\ d^n \frac{|R|(1+|Q|)}{r(r_{\frac{1}{4}} - r_{\frac{1}{2}})(r - r_{\frac{1}{4}})^{n\kappa}} && <\cdot\ d^n \frac{|R|}{r^{n\kappa+2}}, \\
|v_1|_{\frac{3}{4}} &<\cdot\ d^{n^2+n} \frac{|R|(1+|Q|)}{r(r_{\frac{1}{4}} - r_{\frac{1}{2}})(r - r_{\frac{1}{4}})^{n\kappa}(r_{\frac{1}{2}} - r_{\frac{3}{4}})^{n^2\kappa}} && <\cdot\ d^{n^2+n} \frac{|R|}{r^{(n^2+n)\kappa+2}}.
\end{aligned}$$

The factor $(r - r_{\frac{1}{2}})$ in the denominator of the estimates of δ and u is due to estimating the derivative of q_1 with respect to y, and the factor r in the denominators of estimates of B and v_1 is due to the fact that g_1 is the derivative of $T_d R$ with respect to y, evaluated at $y = 0$. In the same estimates the factors $(r_{\frac{1}{4}} - r_{\frac{1}{2}})$ stem from derivatives of v_0 and q_2, respectively. Finally note that the relation $r_\vartheta - r_{\vartheta'} <\cdot\ r$ has been used repeatedly, for $\vartheta - \vartheta' \geq \frac{1}{4}$.

The estimates can be combined in

$$|Y|_{\frac{3}{4}} <\cdot\ d^{n^2+n} \frac{|R|}{r^{(n^2+n)\kappa+1}}, \qquad |\Delta|_{\frac{1}{2}} <\cdot\ d^n \frac{|R|}{r^{n\kappa+1}}. \tag{61}$$

4.4.8. *Mapping of domains.* The following result is needed in the estimates below.

Lemma 4.5. *There is a $C_0 > 0$ such that if the constant C in (43) satisfies $C > C_0$, then*

$$|\Psi - \mathrm{id}|_{\frac{7}{8}}, |\Psi^{-1} - \mathrm{id}|_{\frac{7}{8}} \leq \frac{1-c}{2}(r_{\frac{7}{8}} - r_+) r. \tag{62}$$

In particular, $\Psi(\mathscr{D}_+) \subset \mathscr{D}_{\frac{7}{8}}$ and $\Psi(\mathscr{D}_{\frac{7}{8}}) \subset \mathscr{D}_{\frac{3}{4}}$.

Proof. The first inequality of (61) reads as $|Y|_{\frac{3}{4}} \le C_1 d^{n^2+n}|R|/r^{(n^2+n)\kappa+1}$, where C_1 does not depend on j. Set $C_0 = 4C_1(1-c)/(1-a_1)$. Since $r = (r - r_+)/(1-a_1)$, using induction assumption (43) yields

$$|Y|_{\frac{3}{4}} \le C_1 d^{n^2+n} \frac{|R|}{r^{(n^2+n)\kappa+1}} \le \frac{C_1}{C} r^2 < \frac{C_0}{C}\frac{1}{8}(r-r_+)r < \frac{1-c}{2}(r_{\frac{7}{8}} - r_+)r.$$

For small values of t

$$|\exp(-tY) - \mathrm{id}|_{\frac{7}{8}} = \left| \int_0^t Y \circ \exp(-sY)\,\mathrm{d}s \right|_{\frac{7}{8}} \le t|Y|_{\frac{3}{4}}.$$

Take $z \in \mathscr{D}_{\frac{7}{8}}$. The largest value of t such that $\exp(-tY)(z)$ is still contained in $\mathscr{D}_{\frac{3}{4}}$ is at least equal to 1, since $|Y|_{\frac{3}{4}} < r_{\frac{7}{8}} - r_+ = r_{\frac{3}{4}} - r_{\frac{7}{8}}$. This implies

$$|\Psi - \mathrm{id}|_{\frac{7}{8}}, |\Psi^{-1} - \mathrm{id}|_{\frac{7}{8}} = |\exp(\pm Y) - \mathrm{id}|_{\frac{7}{8}} \le (r_{\frac{7}{8}} - r_+)r, \tag{63}$$

which in turn implies (62). The inclusions follow from this and the fact that $0 < r < 1$. □

4.4.9. *The remainder.* Estimates are needed for $|\check{R}|_+$ and $|\check{Q}|_+$. Recall that

$$\check{R} = [R + \Delta, Y] + (R + \Delta + N[Q,Y] - T_d(R + \Delta + N[Q,Y])) + NS,$$

with $S = \int_0^1 (1-s)[[X+\Delta, Y], Y] \circ \exp(sY)\,\mathrm{d}s$, and

$$\check{Q} = Q + [Q,Y] - N[Q,Y] + S - NS.$$

First, using (61), hypothesis (43) on Q, Cauchy's estimate of derivatives of analytic functions, and Taylor's formula:

$$|[R+\Delta, Y]|_+ <\cdot \frac{|R+\Delta|_{\frac{1}{2}}|Y|_{\frac{3}{4}}}{r_{\frac{3}{4}} - r_+} <\cdot \frac{d^{n^2+2n}}{r^{(n^2+2n)\kappa+3}}|R|^2, \tag{64}$$

$$|[Q,Y] - N[Q,Y]|_+ <\cdot \frac{|Q||Y|_{\frac{3}{4}}}{r_{\frac{3}{4}} - r_+} <\cdot \frac{d^{n^2+n}}{r^{(n^2+n)\kappa+2}}|R|. \tag{65}$$

The terms S and NS are estimated in the same way, using (43) and additionally lemma 4.5:

$$\begin{aligned} |S|_+ &<\cdot \int_0^1 (1-s)\Big|[[X+\Delta,Y],Y]\Big|_{\frac{7}{8}}\,\mathrm{d}s \\ &<\cdot (r_{\frac{3}{4}} - r_{\frac{7}{8}})^{-2}\left(1 + \frac{d^n}{r^{n\kappa+1}}|R|\right)\frac{d^{2n^2+2n}}{r^{(2n^2+2n)\kappa+2}}|R|^2 <\cdot \frac{d^{2n^2+3n}}{r^{(2n^2+3n)\kappa+3}}|R|^2, \end{aligned} \tag{66}$$

$$|NS|_+ <\cdot \frac{d^{2n^2+3n}}{r^{(2n^2+3n)\kappa+3}}|R|^2. \tag{67}$$

Estimating the decay of Fourier coefficients of analytic functions $f : \mathscr{D} \to \mathbb{C}$ with the Paley–Wiener estimate $|f_k| \le |f| \,\mathrm{e}^{-r|k|}$, we obtain:

$$\begin{aligned}
&\Big| R + \Delta + N[Q,Y] - T_d(R + \Delta + N[Q,Y]) \Big|_+ \\
&\quad = \Bigg| \sum_{|k|>d} (R_k + \Delta_k + N[Q,Y]_k)\, \mathrm{e}^{\mathrm{i}\langle k,x\rangle} \Bigg|_+ \\
&\quad \le \sum_{|k|>d} \left(|R| + |\Delta|_{\frac{1}{2}} + \frac{|Q||Y|_{\frac{3}{4}}}{r_{\frac{3}{4}} - r_{\frac{7}{8}}} \right) \mathrm{e}^{-(r_{\frac{7}{8}} - r_+)|k|} \\
&\quad <\cdot\ \frac{d^{n^2+n}}{r^{(n^2+n)\kappa+2}} |R| \int_d^\infty \mathrm{e}^{-(r_{\frac{7}{8}} - r_+)\xi} \xi^{m-1} \,\mathrm{d}\xi \\
&\quad <\cdot\ \frac{d^{n^2+n}}{r^{(n^2+n)\kappa+2}} |R| \frac{\mathrm{e}^{-d(r_{\frac{7}{8}} - r_+)}}{r_{\frac{7}{8}} - r_+} \int_0^\infty \mathrm{e}^{-t} \left(d + \frac{t}{r_{\frac{7}{8}} - r_+} \right)^{m-1} \mathrm{d}t \\
&\quad <\cdot\ |R| \frac{d^{n^2+n+m-1}\, \mathrm{e}^{-d(r_{\frac{7}{8}} - r_+)}}{r^{(n^2+n)\kappa+3}} \left(1 + \left(d(r_{\frac{7}{8}} - r_+) \right)^{-(m-1)} \right); \qquad (68)
\end{aligned}$$

in the last estimate the inequality $(a+b)^n <\cdot\ a^n + b^n$ has been used.

At this point we make an assumption on the growth rates of the geometric sequences $\{r_j\}$ and $\{\rho_j\}$. We require that:

$$0 < a_2 < a_1^{\kappa+1} < 1. \qquad (69)$$

This is equivalent to requiring $d(r_{\frac{7}{8}} - r_+) \sim d\, r \to \infty$ as $j \to \infty$. Under this assumption, combining inequalities (64)-(68) yields:

$$|\check{R}|_+ <\cdot\ \frac{d^{2n^2+3n}}{r^{(2n^2+3n)\kappa+3}} |R| \left(|R| + r^{(n^2+2n)\kappa} d^{m-1-n^2-2n}\, \mathrm{e}^{-rd(1-a_1)/8} \right), \qquad (70)$$

$$|\check{Q} - Q|_+ <\cdot\ \frac{d^{2n^2+3n}}{r^{(2n^2+3n)\kappa+3}} |R|. \qquad (71)$$

4.4.10. *Determining the new vector field.* The next step is to determine the vector field $X_+ = L + R_+ + Q_+$, and to give estimates for R_+ and Q_+. We set $\Phi_+ = \Phi \circ \Psi$, and define:

$$X_+ = \left(\Phi_+^{-1}\right)_* (\tilde{X}_+ + \tilde{\Delta}).$$

Since $\check{X} = \left(\Phi_+^{-1}\right)_* (\tilde{X} + \tilde{\Delta})$, the difference $X_+ - \check{X}$ equals $\left(\Phi_+^{-1}\right)_* (\tilde{X}_+ - \tilde{X})$.

Recall that $\tilde{X}_+ - \tilde{X}$ and $X_+ - \check{X}$ are defined on $\tilde{\mathscr{D}}_+ = (\Phi_{p+}(\mathscr{U}) + \tilde{r}_+) \times \mathcal{P}(\tilde{\rho}_+)$ and $\mathscr{D}_+ = (\mathscr{U} + r_+) \times \mathcal{P}(\rho_+)$ respectively, and that $\Phi_{p+}(\mathscr{U} + r_+) \subset \Phi_{p+}(\mathscr{U}) + \tilde{r}_+ - cr_+$, by virtue of lemma 4.4. The new 'remainder' R_+ equals by definition

$$R_+ = \mathsf{N}X_+ - L = (\mathsf{N}X_+ - \mathsf{N}\check{X}) + (\mathsf{N}\check{X} - L) = \mathsf{N}(\Phi_+^{-1})_*(\tilde{X}_+ - \tilde{X}) + \check{R},$$

and using (70) we have

$$\begin{aligned}
|R_+|_+ &<\cdot\ \frac{\tilde{r}_+}{r_+} |\tilde{X}_+ - \tilde{X}|_{\tilde{\mathscr{D}}_+} + |\check{R}|_+ \\
&<\cdot\ |\tilde{X}_+ - \tilde{X}|_{\tilde{\mathscr{D}}_+} \\
&\qquad + \frac{d^{2n^2+3n}}{r^{(2n^2+3n)\kappa+3}} |R| \left(|R| + r^{(n^2+2n)\kappa} d^{m-1-n^2-2n}\, \mathrm{e}^{-rd(1-a_1)/8} \right), \qquad (72)
\end{aligned}$$

Likewise, using (71), we find that

$$|Q_+|_+ <\cdot\ |Q| + |\tilde{X}_+ - \tilde{X}|_{\tilde{\mathscr{D}}_+} + \frac{d^{2n^2+3n}}{r^{(2n^2+3n)\kappa+3}}|R|. \tag{73}$$

4.4.11. *The induction hypothesis.* The induction hypothesis of subsection 4.4.1 has to be verified for $j+1$. Note that the geometrically decreasing sequences r_j *etc.* have not yet been fully specified; only a number of conditions — (47), (48), (69) — have been given which they have to satisfy. We give here for every statement in the induction hypothesis sufficient conditions.

Condition (41) is vacuous if $j=1$. We have to show that it holds for $i=j$, if the induction hypothesis is satisfied for $i<j$; that is, we have to show that $|\Psi - \mathrm{id}|_+ < cr_+$. It follows from (48) and (63) that

$$|\Psi - \mathrm{id}|_+ < (r_{\frac{7}{8}} - r_+)r = \frac{r}{8}\left(\frac{1}{a_1} - 1\right) r_+.$$

Therefore (41) is certainly satisfied if

$$\frac{r_0}{8}\left(\frac{1}{a_1} - 1\right) < c; \tag{74}$$

for given c, this condition can always be satisfied if r_0 is chosen sufficiently small.

Condition (42) is for $j=1$ trivially satisfied, since $\Phi_1 = \mathrm{id}$ and φ_1 and ψ_1 vanish identically. For $i=j+1$ the condition can be written as

$$|\varphi_+|_+, |D\varphi_+|_+, |D\psi_+|_+ < c - \frac{r_+}{1-a_1}.$$

Details are given only for the estimate of $D\psi_+$, the others being easier. Note first that

$$\psi_+ = \Phi_+^{-1} - \mathrm{id} = \Psi^{-1}\circ\Phi^{-1} - \mathrm{id} = (\Psi^{-1} - \mathrm{id})\circ\Phi^{-1} + (\Phi^{-1} - \mathrm{id}).$$

By placing the condition $(1+c)r_+ < r_{\frac{15}{16}}$ on c, or equivalently, by demanding that

$$c < \frac{1}{16}\left(\frac{1}{a_1} - 1\right), \tag{75}$$

we ensure that Φ^{-1} maps the domain $\mathscr{D}_+$ inside $\mathscr{D}_{15/16}$; now we can estimate the derivative of $\Psi^{-1} - \mathrm{id}$ on this domain.

$$\begin{aligned}
|D\psi_+|_+ &\le |D(\Psi^{-1} - \mathrm{id})|_{\frac{15}{16}} |D\Phi^{-1}|_+ + |D(\Phi^{-1} - \mathrm{id})|_+ \\
&\le \frac{|\Psi^{-1} - \mathrm{id}|_{\frac{7}{8}}}{r_{\frac{7}{8}} - r_{\frac{15}{16}}}\left(1 + c - \frac{r}{1-a_1}\right) + |D\psi|_+ \\
&\le (1-c)r\left(1 + c - \frac{r}{1-a_1}\right) + c - \frac{r}{1-a_1} \\
&< (1-c^2)r + c - \frac{r_+}{1-a_1} + \frac{r_+ - r}{1-a_1} \\
&< (1-c^2)r + c - \frac{r_+}{1-a_1} - r \\
&< c - \frac{r_+}{1-a_1}.
\end{aligned}$$

The first part of condition (43) is satisfied for $j=1$ if the size ε of the initial perturbation is sufficiently small; the second part can be satisfied by choosing r_0

sufficiently small. To show that these conditions hold for $j+1$, if they hold for j, is the subject of the next subsection.

Excepting this last verification, we have the following conditions on the sequences $\{r_j\}$, $\{\tilde r_j\}$, $\{\rho_j\}$, $\{\tilde\rho_j\}$, $\{d_j\}$ (conditions (47), (48), (69) and (75), together with the condition that $\rho_0 > 0$ is small enough to imply $d_1 > 2$:

$$\begin{aligned}
&\frac{\tilde r_+}{\tilde r} = \frac{r_+}{r} = a_1, \qquad \frac{\tilde\rho_+}{\tilde\rho} = \frac{\rho_+}{\rho} = a_2,\\
&1 + 2c < \frac{\tilde r_0}{r_0} < (1-c)\left(\frac{1}{2a_1} + 1\right),\\
&0 < a_2 < a_1^{\kappa+1} < 1,\\
&0 < c < \frac{1}{16}\left(\frac{1}{a_1} - 1\right).
\end{aligned}$$

If r_0 and a_1 are given, then a_2, c and $\tilde r_0$ can always be found such that these inequalities all hold. Note therefore that we are always free to choose r_0 and a_1, provided $r_0 > 0$ and $0 < a_1 < 1$.

4.5. **Smallness of the remainder term.** The sequences $r_j = r_0 a_1^j$ and $\rho_j = \rho_0 a_2^j$ have now to be determined in such a way that $|R_{j+1}|_{j+1} \ll |R_j|_j$; in the next subsection, this will be shown to ensure that the embeddings $\Phi_j = \Psi_1 \circ \cdots \circ \Psi_{j-1}$ converge to an embedding Φ_∞ that has the properties stated in theorem 2.3. Note that from this point onwards, the '+'- and '$\cdot$'-notations are dropped.

Inequality (72) reads then as

$$\begin{aligned}
|R_{j+1}|_{j+1} < C\,|\tilde X_{j+1} - \tilde X|_{\tilde{\mathcal D}_{j+1}}\\
+ C\frac{d_j^{2n^2+3n}}{r_j^{(2n^2+3n)\kappa+3}}|R_j|_j\left(|R_j|_j + r_j^{(n^2+2n)\kappa} d_j^{m-1-n^2-2n}\,\mathrm{e}^{-r_j d_j(1-a_1)/8}\right).
\end{aligned} \tag{76}$$

where the constant C does not depend on j. Recall that the truncation level is defined in (58), which reads as

$$d_j = \frac{1}{2}\left(\frac{\gamma}{2\gamma_0\rho_0}\right)^{\frac{1}{\kappa+1}}\left(a_2^{\frac{1}{\kappa+1}}\right)^j. \tag{77}$$

We introduce $\varepsilon = \|P\|_{\mathcal B}$. There are several cases, depending on the smoothness class of the original perturbed vector field $\tilde X = X + P$. If $\tilde X$ is real analytic, then $\tilde X_j = \tilde X$ for all j, and the first term in (76) vanishes. If $\tilde X$ fails to be real analytic, there is an approximating holomorphic sequence $\tilde X_j$ satisfying

$$|\tilde X_{j+1} - \tilde X_j|_{\tilde{\mathcal D}_{j+1}} \le \varepsilon b_{j+1},$$

where the b_j are given by lemmas 4.1 or 4.3. In particular, if $\tilde X$ is Gevrey regular, approximations can be found for which the quantity $\log 1/b_j$ increases exponentially in j. If $\tilde X$ is not Gevrey, but still in some Carleman class, then $\log 1/b_j$ increases slower than exponentially, but faster than any linear function in j. Finally, in the finitely differentiable class, the sequence $\log 1/b_j$ increases linearly with j.

For each of these four cases, a sequence $\{\delta_j\}$ will be determined that decreases monotonically towards 0, such that, under appropriate conditions,

$$|R_j|_j \le \delta_j \quad \text{for all } j \in \mathbb{N}. \tag{78}$$

First, we make some definitions that will hold for several of the cases considered below. If $a_1 \in (0,1)$ is fixed, we choose $a_2 \in (0, a_1^{\kappa+1})$ such that

$$1 < \beta \stackrel{\text{def}}{=} \frac{a_1}{a_2^{\frac{1}{\kappa+1}}} < 2.$$

With this choice, and setting

$$c_0 = \frac{1}{8} r_0 d_0 (1-a_1) = \frac{1}{16}(1-a_1)\left(\frac{\gamma}{2\gamma_0}\right)^{\frac{1}{\kappa+1}} \frac{r_0}{\rho_0^{1/(\kappa+1)}}, \tag{79}$$

we have that

$$\mathrm{e}^{-r_j d_j (1-a_1)/8} = \exp\left(-c_0 a_1^j a_2^{-j/(\kappa+1)}\right) = \exp\left(-c_0 \beta^j\right).$$

4.5.1. *Case one: real analyticity.*

Lemma 4.6. *Let $\tilde{X} \in \mathcal{X}_h^\omega$ be real analytic. If $\varepsilon_0 > 0$, $r_0 > 0$ and $\rho_0 > 0$ are sufficiently small, and if $\delta_j = \varepsilon\,\mathrm{e}^{-\beta^j}$ for $0 < \varepsilon < \varepsilon_0$, then* (78) *holds for all j.*

Proof. Recall that $0 < \varepsilon < \varepsilon_0$. We proceed by induction. It is given that $|R_0|_0 \le \varepsilon$. With the induction assumption $|R_j|_j \le \varepsilon\,\mathrm{e}^{-\beta^j}$, inequality (76) reads as

$$\frac{|R_{j+1}|_{j+1}}{\varepsilon\,\mathrm{e}^{-\beta^{j+1}}} < C \frac{d_j^{2n^2+3n}}{r_j^{(2n^2+3n)\kappa+3}} \varepsilon\,\mathrm{e}^{(-2+\beta)\beta^j} + C \frac{d_j^{n^2+n+m-1}}{r_j^{(n^2+n)\kappa+3}} \mathrm{e}^{\beta^j\left(-1-c_0+\beta\right)}, \tag{80}$$

where $c_0 = r_0 d_0 (1-a_1)/8$. For given sequences d_j and r_j, the first term in this sum can be made smaller than $1/2$ by choosing $\varepsilon_0 > 0$ sufficiently small.

The second term is of the form $\mathrm{e}^{f(j)}$, where $f(x) = \log C_0 + x \log\alpha - A\beta^x$, with $A = c_0 + 1 - \beta$ and α only depending on a_1, a_2, κ, m and n, but not on r_0 and ρ_0. Computing f', we see that this concave function, restricted to $x \ge 0$, takes its maximum at

$$x_* = \frac{1}{\log\beta} \log\left(\frac{1}{A}\frac{\log\alpha}{\log\beta}\right),$$

if $\log\alpha/\log\beta \ge A$, otherwise at $x_* = 0$.

If we take ρ_0 sufficiently small, and, by (79), consequently c_0 and A sufficiently large, the second case occurs; the value of the maximum is then $f(0) = \log C_0 - A$. It follows that

$$\frac{d_j^{n^2+n+m-1}}{r_j^{(n^2+n)\kappa+3}} \mathrm{e}^{\beta^j\left(-1-c_0+\beta\right)} \le \frac{d_0^{n^2+n+m-1}}{r_0^{(n^2+n)\kappa+3}} \mathrm{e}^{-c_0-1+\beta}.$$

Note that by fixing r_0 and taking ρ_0 sufficiently small, again by invoking (79) the right hand side can be made smaller than $1/2$. It follows that we can make the right hand side of (80) smaller than 1, uniformly in j, by taking $\varepsilon_0 > 0$ and ρ_0 sufficiently small. $\square$

4.5.2. *Case two: Gevrey regularity.* If $\tilde{X}$ is in the Gevrey class $\mathcal{X}_h^\mu$, we can find an holomorphic approximating sequence $\tilde{X}_j$ such that equation (38) holds, that is, such that for some $\eta > 0$

$$|\tilde{X}_{j+1} - \tilde{X}_j|_{\tilde{\mathcal{D}}_{j+1}} < C\varepsilon \exp\left(-C' r_j^{-\frac{1}{\mu+2\eta}}\right).$$

We take $a_1 \in (0,1)$ and $a_2 \in (0, a_1^{\kappa+1})$ such that $\beta = a_1^{-1/(\mu+2\eta)} = a_1/a_2^{1/(\kappa+1)} < 2$.

Lemma 4.7. *Let $\tilde{X}$ be in the Gevrey class $\mathcal{X}_h^\mu$, take $\eta > 0$ and $\delta_j = \varepsilon\, \mathrm{e}^{-\beta^j}$ with $0 < \varepsilon < \varepsilon_0$. If $\varepsilon > 0$, $r_0 > 0$ and $\rho_0 > 0$ are sufficiently small, then* (78) *holds.*

Proof. The proof resembles that of the previous lemma. Using (38), inequality (76) reads as

$$\frac{|R_{j+1}|_{j+1}}{\varepsilon\, \mathrm{e}^{-\beta^{j+1}}} < C\, \mathrm{e}^{-\beta^j\left(C' r_0^{-\frac{1}{\mu+2\eta}} - \beta\right)} + C\frac{d_j^{2n^2+3n}}{r_j^{(2n^2+3n)\kappa+3}}\varepsilon\, \mathrm{e}^{(-2+\beta)\beta^j} + C\frac{d_j^{n^2+n+m-1}}{r_j^{(n^2+n)\kappa+3}}\, \mathrm{e}^{\beta^j\left(-1-c_0+\beta\right)},$$

where $c_0 = r_0 d_0(1-a_1)/8$. The first term can be made smaller than $1/3$ by taking r_0 sufficiently small. It follows exactly as in the proof of lemma 4.6 that if $\varepsilon > 0$, $r_0 > 0$ and $\rho_0 > 0$ are sufficiently small, the other two terms are both smaller than $1/3$, making the right hand side is smaller than 1, uniformly in j. $\square$

4.5.3. *Case three: ultradifferentiability.* If $\tilde{X}$ is in the Carleman class $\mathcal{X}^M$, that is, if it is infinitely differentiable but not Gevrey regular, let $\{\lambda_s\}$ be the sequence given in (37), and let $\lambda_* : [0,\infty) \to \mathbb{R}$ be its largest convex minorant.

We construct a function g_M as follows. Let $g_0 = \mathcal{L}\lambda_*(0)$ and let

$$g_j = \min\Big\{\beta g_{j-1}, \mathcal{L}\lambda_*\left(\log r_j^{-1}\right)\Big\}. \tag{81}$$

Finally, let g_M be the convex function whose epigraph equals the convex hull of the points $(\log r_j^{-1}, g_j)$ and the half-line $\{(0, g_0 + t) \,|\, t \geq 0\}$. Then g_M is a convex minorant of $\mathcal{L}\lambda_*$, which moreover satisfies

$$g_M(\log r_{j+1}^{-1}) \leq \bar{\beta} g_M(\log r_j^{-1}).$$

Since g_M is a minorant of $\mathcal{L}\lambda_*$, it follows from lemma 4.3 that there is an approximating sequence $\tilde{X}_j$ such that

$$|\tilde{X}_{j+1} - \tilde{X}_j|_{\tilde{\mathcal{D}}_{j+1}} < \varepsilon\, \mathrm{e}^{-g_M(\log r_j^{-1})}.$$

The sequence $\{\sigma_j\}$ given by $\sigma_j = g_M(\log r_{j-1}^{-1})$ has by construction of g_M the property that $\sigma_{j+1} < \beta\sigma_j$ for all j. Note that it follows from lemma 4.3 that σ_j increases faster than any linear function of j.

Lemma 4.8. *Let $\tilde{X} \in \mathcal{X}_h^M$, $\eta > 0$ and set $\delta_j = C_2\varepsilon\, \mathrm{e}^{-\sigma_j}$, where $0 < \varepsilon < \varepsilon_0$. If $C_2 > 0$ is chosen sufficiently large, and $\varepsilon_0 > 0$, $r_0 > 0$ and $\rho_0 > 0$ are small, then* (78) *holds.*

Proof. As before; inequality (76) reads as

$$\frac{|R_{j+1}|_{j+1}}{C_2\varepsilon\, \mathrm{e}^{-\sigma_{j+1}}} < C\frac{|\tilde{X}_{j+1} - \tilde{X}_j|_{\tilde{\mathcal{D}}_{j+1}}}{C_2\varepsilon\, \mathrm{e}^{-\sigma_{j+1}}} + CC_2\frac{d_j^{2n^2+3n}}{r_j^{(2n^2+3n)\kappa+3}}\varepsilon\, \mathrm{e}^{\left(\sigma_{j+1}-2\sigma_j\right)} + C\frac{d_j^{n^2+n+m-1}}{r_j^{(n^2+n)\kappa+3}}\, \mathrm{e}^{\left(\sigma_{j+1}-\sigma_j-c_0\beta^j\right)}.$$

If we choose $C_2 = 3C$, the first term is at most equal to $1/3$. By the choice of the σ_j, we have $\sigma_{j+1} - 2\sigma_j < (\bar{\beta} - 2)\sigma_j < 0$; as a consequence, the second term on the right hand side can be made smaller that $1/3$, uniformly in j, if $\varepsilon_0 > 0$ is taken sufficiently small. Moreover, since $\sigma_j \leq \sigma_0 \beta^j$, we have that

$$\sigma_{j+1} - \sigma_j - c_0\beta^j \leq c_0\beta^j \left(\frac{(\beta - 1)\sigma_0}{c_0} - 1 \right) \leq -\frac{c_0}{2}\beta^j, \tag{82}$$

where the last inequality follows from taking ρ_0 sufficiently small, hence c_0 sufficiently large. The third term can now be made smaller than $1/3$ by choosing ρ_0 sufficiently small, thereby making c_0 as large as is required. □

4.5.4. *Case four: finite differentiability.* In the case that $\tilde{X} \in \mathcal{X}^s$, we obtain from lemma 4.1 that

$$|\tilde{X}_{j+1} - \tilde{X}_j|_{\tilde{\mathcal{D}}_{j+1}} < C_s r_j^s \varepsilon.$$

If $s > N \overset{\text{def}}{=} (2n^2 + 3n)(\kappa + 1) + 3$, then $a_1^s < a_1^N$. For $s > N$, we take $a_2 < a_1^{\kappa+1}$ so close to $a_1^{\kappa+1}$ such that the interval

$$I = \left(a_1^s, a_1^{(2n^2+3n)\kappa+3} a_2^{\frac{2n^2+3n}{\kappa+1}} \right) \subset \left(a_1^s, a_1^N \right)$$

is not empty, and we choose $a_3 \in I$.

Lemma 4.9. *Let $\tilde{X} \in \mathcal{X}^s$, take $\eta > 0$ and $\delta_j = \varepsilon a_3^j$, with $0 < \varepsilon < \varepsilon_0$. If $\varepsilon_0 > 0$, $r_0 > 0$ and $\rho_0 > 0$ are sufficiently small, then (78) holds.*

Proof. As before; equation (76) can in this final case be written as

$$\frac{|R_{j+1}|_{j+1}}{\varepsilon a_3^{j+1}} < C \left(r_0^s \left(\frac{a_1^s}{a_3} \right)^{j+1} + \varepsilon \frac{d_0^{2n^2+3n}}{a_3 r_0^{(2n^2+3n)\kappa+3}} \left(\frac{a_3}{a_1^{(2n^2+3n)\kappa+3} a_2^{(2n^2+3n)/(\kappa+1)}} \right)^j \right.$$
$$\left. + \frac{d_0^{n^2+n+m-1}}{a_3 r_0^{(n^2+n)\kappa+3}} \left(a_1^{(n^2+n)\kappa+3} a_2^{\frac{n^2+n+m-1}{\kappa+1}} a_3 \right)^{-j} \mathrm{e}^{-c_0\beta^j} \right).$$

The first two terms on the right hand side are decreasing geometrical series, which can each be made smaller than $1/3$, uniformly in j, by choosing $\varepsilon_0 > 0$ and $r_0 > 0$ sufficiently small. The third term can be made smaller than $1/3$ by choosing $\rho_0 > 0$ sufficiently small, thereby making c_0 as large as is required. □

4.6. **Convergence.** Let $\mathscr{D}_\infty = \bigcap_{j=1}^\infty \mathscr{D}_j$.

Lemma 4.10. *For $\tilde{X}$ in one of the four smoothness classes $\mathcal{X}^\omega$, $\mathcal{X}^\mu$, $\mathcal{X}^M$ and $\mathcal{X}^s$, where $s > (2n^2+3n)(\kappa+1)+3$, let the hypotheses regarding the smallness of $\varepsilon_0, r_0, \rho_0$ and $1/C_2$ of the corresponding lemma 4.6-4.9 be fulfilled. Then there is a conjugacy $\Phi_\infty : \mathscr{D}_\infty \to \mathscr{M} \times \mathscr{P}$, such that*

$$\Phi_j \to \Phi_\infty \quad \text{as } j \to \infty,$$

together with at least its derivatives up to order smaller than $(n^2+n)\kappa+2$ with respect to the phase variables, uniformly on $\mathscr{D}_\infty$; we have that $\|\Phi_\infty - \mathrm{id}\|_{\mathcal{B}'} \leq C\|P\|_{\mathcal{B}}$. Moreover,

$$\Lambda_j \to \Lambda_\infty \quad \text{as } j \to \infty,$$

$\|\Lambda_\infty\|_{\mathcal{B}'} \leq C\|P\|_{\mathcal{B}}$ and $\mathsf{N}\Phi_{\infty*}^{-1}(\tilde{X} + \Lambda_\infty) = L$. Additionally, we have the following. Let $\zeta > 0$ be a fixed constant.

1. *If $\tilde{X} \in \mathcal{X}_h^\omega$ is real analytic, then Φ_∞ is real analytic in the phase variables and Φ_∞ and Λ_∞ are Gevrey $\mathcal{G}_h^\nu$-regular in the parameters, where*
$$\nu = 1 + \ell(\kappa+1) + \zeta.$$
2. *If $\tilde{X} \in \mathcal{X}_h^\mu$ is Gevrey regular, then Φ_∞ is $\mathcal{G}^{\nu_1}$-regular in the phase variables and Φ_∞ and Λ_∞ are $\mathcal{G}^{\nu_2}$-regular in the parameters, where*
$$\nu_1 = 1 + \mu + \zeta, \quad \nu_2 = 1 + \ell\mu(\kappa+1) + \zeta.$$
3. *If $\tilde{X} \in \mathcal{X}_h^M$, let $D^\alpha = D_{(x,y)}^{\alpha_1} D_p^{\alpha_2}$. Then there are constants $C_1, C_2 > 0$ such that*
$$\sup_{\mathscr{D}_\infty} |D^\alpha \Phi_\infty| \le C_1^{1+|\alpha|} \alpha! \, \mathrm{e}^{\mathcal{L} g_M(|\alpha_1| + \ell(\kappa+1+\zeta)|\alpha_2| + C_2)}.$$
An analogous estimate holds for Λ_∞.
4. *If $\tilde{X} \in \mathcal{X}^s$, then the conjugacy Φ_∞ is $\mathcal{C}^{s-(n^2+n)\kappa-2-\zeta}$ in the phase direction and $\mathcal{C}^{(s-(n^2+n)\kappa-2-\zeta)/(\ell(\kappa+1))}$ in the parameter direction; the modifying terms vector field Λ_∞ is $\mathcal{C}^{(s-n(\kappa+1)-2-\zeta)/(\ell(\kappa+1))}$ in the parameters.*

   In particular, the conjugacy is at least $\mathcal{C}^{(n^2+2n)(\kappa+1)+n^2+n+1}$ in the phase direction and $\mathcal{C}^{n+1}$ in the parameter direction; the modifying terms vector field is at least $\mathcal{C}^{2(n+1)}$ in the parameters.

Proof. In this proof, we use a series of constants $C_1, C_2, \cdots$ that are unrelated to any constants of the same name used earlier. On the domains $\mathscr{D}_{j+1}$, the following estimates obtain:
$$\begin{aligned} |\Phi_{j+1} - \Phi_j|_{j+1} &= \left|\Phi_j \circ \Psi_j - \Phi_j \circ \mathrm{id}_{\mathscr{D}_{j+1}}\right|_{j+1} \\ &\le |D\Phi_j|_{j+\frac{3}{4}} \left|\mathrm{e}^{Y_j} - \mathrm{id}_{\mathscr{D}_{j+1}}\right|_{j+1} \\ &\le C_1 \frac{|\Phi_j|_j}{r_{j+1} - r_j} \frac{d_j^{n^2+n}}{r_j^{(n^2+n)\kappa+1}} \delta_j \le C_2 \frac{d_j^{n^2+n}}{r_j^{(n^2+n)\kappa+2}} \delta_j |\Phi_j|_j. \end{aligned} \tag{83}$$
In the second inequality, we have used (61). From this, it follows that
$$|\Phi_{j+1}|_{j+1} \le |\Phi_{j+1} - \Phi_j|_{j+1} + |\Phi_j|_j \le \left(1 + C_2 \frac{d_j^{n^2+n}}{r_j^{(n^2+n)\kappa+2}} \delta_j\right) |\Phi_j|_j,$$
and consequently that
$$|\Phi_j|_j \le \prod_{i=1}^{j-1} \left(1 + C_2 \frac{d_i^{n^2+n}}{r_i^{(n^2+n)\kappa+2}} \delta_i\right) |\Phi_1|_1.$$
Since $\Phi_1 = \mathrm{id}$, and as a consequence of lemmas 4.6-4.9, for every smoothness class the product on the right hand side can be bounded by some constant C_3, uniformly in j. Inequality (83) then implies that
$$|\Phi_{j+1} - \Phi_j|_{j+1} \le C_4 \frac{d_j^{n^2+n}}{r_j^{(n^2+n)\kappa+2}} \delta_j. \tag{84}$$
From this and the form of δ_j given in the respective lemma 4.6-4.9 it follows that the infinite sum on the right hand side of
$$\Phi_\infty - \mathrm{id} = \Phi_1 - \mathrm{id} + \sum_{j=1}^{\infty} (\Phi_{j+1} - \Phi_j) = \sum_{j=1}^{\infty} (\Phi_{j+1} - \Phi_j)$$

converges absolutely and uniformly on the intersection $\mathscr{D}_\infty = \bigcap_{j=1}^\infty \mathscr{D}_j$, and the limit Φ_∞ is therefore at least continuous there.

Let $\alpha = (\alpha_1, \alpha_2) = (\alpha_1^x, \alpha_1^y, \alpha_2) \in \mathbb{N}^m \times \mathbb{N}^n \times \mathbb{N}^q$ be a multi-index, and let $D^\alpha = D_x^{\alpha_1^x} D_y^{\alpha_1^y} D_p^{\alpha_2}$. The derivative $D^\alpha \Phi_\infty$ will exist on $\mathscr{D}_\infty$, in the sense of Whitney, if the series

$$D^\alpha(\Phi_\infty - \mathrm{id}) = \sum_{j=1}^\infty (D^\alpha \Phi_{j+1} - D^\alpha \Phi_j) \tag{85}$$

converges uniformly on $\mathcal{D}_\infty$. To see this, take

$$\begin{aligned}|D^\alpha \Phi_j - D^\alpha \Phi_{j-1}|_{j+1} &\le C_5^{\ell|\alpha_2|} \alpha! \frac{|\Phi_j - \Phi_{j-1}|_j}{(r_j - r_{j+1})^{|\alpha_1|} (\rho_j^\ell - \rho_{j+1}^\ell)^{|\alpha_2|}} \\ &\le C_6^{|\alpha|} \alpha! \frac{\delta_j}{r_j^{|\alpha_1| + (n^2+n)\kappa + 2} \rho_j^{|\alpha_2| \ell}}.\end{aligned}$$

In the first inequality, we used (25). Set

$$A = (|\alpha_1| + (n^2+n)\kappa + 2) \log a_1^{-1} + \ell |\alpha_2| \log a_2^{-1},$$

then $A > 0$ and

$$|D^\alpha \Phi_j - D^\alpha \Phi_{j-1}|_{j+1} \le C_7^{|\alpha|} \alpha! \, \mathrm{e}^{\log \delta_j + Aj} \quad \text{for all } j. \tag{86}$$

4.6.1. *Finite differentiability.* In the finitely differentiable case, $\delta_j = \varepsilon a_3^j$, and the right hand side of this inequality is a decreasing geometric series if and only if $\log a_3^{-1} > A$. Choosing $a_3 = a_1^{s-\zeta}$, and $a_2 = a_1^{\kappa+1+\zeta}$, for some $\zeta > 0$, this condition reads as

$$|\alpha_1| + (n^2+n)\kappa + 2 + \ell|\alpha_2|(\kappa + 1 + \zeta) \le s - \zeta.$$

As $\zeta > 0$ was arbitrary, this is implied by

$$|\alpha_1| + \ell(\kappa+1)|\alpha_2| < s - (n^2+n)\kappa - 2.$$

This inequality describes exactly the anisotropic differentiability (in the sense of [28]) of the conjugacy in the presence of a multiple normal eigenvalue of multiplicity ℓ. We find that for all α satisfying this inequality that

$$|D^\alpha(\Phi_\infty - \mathrm{id})|_\infty \le C_\alpha \varepsilon.$$

4.6.2. *A lemma.* We need the following result a couple of times.

Lemma 4.11. *Let $g : [0,\infty) \to \mathbb{R}$ be an increasing convex function, and let $f = \mathcal{L}g$ be its Legendre transformation. Then*

$$\sum_{j=1}^\infty \mathrm{e}^{-g(j)} \le \frac{\mathrm{e}^{f(1)}}{\mathrm{e} - 1}.$$

Proof. By definition of the Legendre transformation,

$$g(p) = \max_x \{px - f(x)\};$$

in particular, taking $x = 1$ and $p = j$,

$$g(j) \ge j - f(1).$$

The result follows from this. $\square$

4.6.3. *Real analyticity and Gevrey regularity.* In the real analytic and the Gevrey cases (see lemmas 4.6 and 4.7), we have that $a_2 = (a_1/\beta)^{\kappa+1}$. Put $g(p) = \beta^p - Ap + \log \varepsilon^{-1}$ and note that $g(j) = -(\log \delta_j + Aj)$ for all $j \in \mathbb{N}$. Remark that g is a convex function, and using (36) it follows that its Legendre transform $f = \mathcal{L}g$ is equal to

$$f(x) = \frac{x+A}{\log \beta} \left(\log \frac{x+A}{\log \beta} - 1 \right) + \log \varepsilon. \tag{87}$$

Using lemma 4.11, it follows that

$$\sum_{j=1}^{\infty} |D^\alpha \Phi_j - D^\alpha \Phi_{j-1}|_{j+1} \le C_7^{|\alpha|} \alpha! \sum_{j=1}^{\infty} \mathrm{e}^{-g(j)} < C_8^{|\alpha|} \alpha! \, \mathrm{e}^{f(1)}. \tag{88}$$

With equation (87), this yields that

$$|D^\alpha(\Phi_\infty - \mathrm{id})|_\infty < \varepsilon \, C_9^{|\alpha|} \left(|\alpha_1| + |\alpha_2|\right)^{|\alpha_1| \left(1 + \frac{\log a_1^{-1}}{\log \beta}\right) + |\alpha_2| \left(1 + \frac{\ell \log a_2^{-1}}{\log \beta}\right)}. \tag{89}$$

In the analytic case, the domain $\mathscr{D}_\infty$ contains an open complex strip around the real phase space, so the inference that Φ_∞ is real analytic in the phase directions follows directly from the boundedness of $|\Phi_\infty|_\infty$. For the regularity of the parameter dependence, take $\zeta > 0$ sufficiently close to 0 and set $a_1 = \mathrm{e}^{-\zeta}$. Then

$$1 + \frac{\ell \log a_2^{-1}}{\log \beta} = 1 + \frac{\ell \log a_2^{-1}}{-\zeta + \frac{1}{\kappa+1} \log a_2^{-1}} < 1 + \ell(\kappa+1) + \eta,$$

and we see Φ_∞ is $\mathcal{G}^\nu$-regular in the parameter direction if $\nu > 1 + \ell(\kappa+1)$. Note that this generalises the result of Popov [25, 26] to the case of multiple normal eigenvalues.

In the Gevrey case, we have that $\log \beta = \log a_1^{-1}/(\mu + \eta)$. From (89) we infer that Φ_∞, as well as all its derivatives in the parameter direction, is $\mathcal{G}^{\nu_1}$ regular in the phase variables, and Φ_∞ together with its derivatives in the phase variables is $\mathcal{G}^{\nu_2}$-regular in the parameters, with

$$\nu_1 > 1 + \mu, \quad \text{and} \quad \nu_2 > 1 + \mu\ell(\kappa+1).$$

4.6.4. *Ultradifferentiability.* The final case occurs if $\tilde{X}$ is ultradifferentiable, but not Gevrey regular. Equation (86) reads then as

$$|D^\alpha \Phi_t - D^\alpha \Phi_{j-1}|_{j+1} \le \varepsilon C_2 C_7^{|\alpha|} \alpha! \, \mathrm{e}^{-\sigma_j + Aj} \quad \text{for all } j. \tag{90}$$

Recall that

$$\sigma_j = g_M(-\log r_{j-1}) = g_M\Big((j-1)\log a_1^{-1} + \log r_0^{-1}\Big)$$

for all j. Introducing

$$g(p) = g_M(p \log a_1^{-1} + \log r_0^{-1}) - A(p+1),$$

the estimate of equation (90) can be written as

$$|D^\alpha \Phi_t - D^\alpha \Phi_{j-1}|_{j+1} \le \varepsilon C_2 C_7^{|\alpha|} \alpha! \, \mathrm{e}^{-g(j-1)} \quad \text{for all } j.$$

Analogously to 4.6.3, the convergence of the sum can be expressed in terms of the Legendre transformation $\mathcal{L}g$ of g.

Note that if g is an increasing convex function, the Legendre transformation of

$$h(p) = g(ap+b) - cp - d$$

is equal to

$$\mathcal{L}h(x) = \mathcal{L}g\left(\frac{x+c}{a}\right) - b\frac{x+c}{a} + d.$$

Using this relation with $a = \log a_1^{-1}$, $b = \log r_0^{-1}$, $c = A$ and $d = A$ and recalling that the Legendre transform is involutive (*i.e.* $\mathcal{L}^2 g = g$), we see that

$$\mathcal{L}g(x) = \mathcal{L}g_M\left(\frac{x+A}{\log a_1^{-1}}\right) - \frac{\log r_0^{-1}}{\log a_1^{-1}}(x+A) + A.$$

If $\zeta > 0$ is fixed and if we take $a_2 = a_1^{\kappa+1+\zeta}$, then

$$\mathcal{L}g(1) = \mathcal{L}g_M\Big(|\alpha_1| + \ell(\kappa+1+\zeta)|\alpha_2| + C_{10}\Big).$$

Finally, we obtain

$$|D^\alpha(\Phi_\infty - \mathrm{id})|_\infty < \varepsilon C_2 C_7^{|\alpha|}\alpha!\, \mathrm{e}^{\mathcal{L}g_M(|\alpha_1|+\ell(\kappa+1+\zeta)|\alpha_2|+C_{10})}. \tag{91}$$

4.6.5. *Convergence of the modifying terms.* The estimates for Λ_∞ follow entirely analogously from the fact that $\tilde{\mathscr{D}}_{j+1} \subset \Phi_j(\mathscr{D}_{j+\frac{1}{2}})$, which implies that $\tilde{\Delta}_j = \Phi_{j*}\Delta_j$ is well-defined, and that

$$|\tilde{\Delta}_j|_{\tilde{\mathscr{D}}_{j+1}} \le |\Phi_{j*}\Delta_j|_{j+\frac{1}{2}} \le C_{12}|D\,\Phi_j|_{j+\frac{1}{2}}|\Delta_j|_{j+\frac{1}{2}} \le C_{13}|\Phi_j|_j \frac{d_j^n}{r_j^{n\kappa+2}}\delta_j,$$

which ensures the absolute and uniform convergence of

$$\Lambda_\infty = \sum_{j=1}^{\infty} \tilde{\Delta}_j.$$

Since moreover

$$|D_p^{\alpha_2}\tilde{\Delta}_j|_{\tilde{\mathscr{D}}_{j+1}} \le C_{14}|\Phi_j|_j \frac{d_j^n}{r_j^{n\kappa+2}\rho_j^{|\alpha_2|\ell}}\delta_j,$$

it follows analogously to the proof of the convergence of $D^\alpha\Phi_j$ that

$$D_p^{\alpha_2}\Lambda_\infty = \sum_{j=1}^{\infty} D_p^{\alpha_2}\tilde{\Delta}_j.$$

converges absolutely and uniformly on $\tilde{\mathscr{D}}_\infty = \bigcap_{j=1}^\infty \tilde{\mathscr{D}}_j$. □

This concludes the proof of theorem 2.3.

REFERENCES

[1] V.I. Arnol'd, *Geometrical methods in the theory of ordinary differential equations*, Springer, Heidelberg, 1988, Second Edition.

[2] ———, *Mathematical methods of classical mechanics*, Springer, Heidelberg, 1989.

[3] R.I. Bogdanov, *Bifurcations of a limit cycle of a family of vector fields in the plane*, Trudy Seminara Imeni I.G. Petrovskogo **2** (1976), 23–36, Translation of: Selecta Math. Sovietica, **1**, (1981), 373–388.

[4] ———, *A versal deformation of a singular point of a vector field in the plane in the case of vanishing eigenvalues*, Trudy Seminara Imeni I.G. Petrovskogo **2** (1976), 37–65, Translation of: Selecta Math. Sovietica, **1**, (1981), 389–421.

[5] J. Bonet, R.W. Braun, R. Meise, and B.A. Taylor, *Whitney's extension theorem for nonquasi-analytic classes of ultradifferentiable functions*, Studia Mathematica **99** (1991), 155–184.

[6] H.W. Broer, H. Hanßmann, and J. You, *Bifurcations of normally parabolic tori in Hamiltonian systems*, Nonlinearity **18** (2005), 1735–1769.

[7] H.W. Broer, J. Hoo, and V. Naudot, *Normal linear stability of quasi-periodic tori*, Journal of Differential Equations **232** (2007), 355–418.

[8] H.W. Broer, G.B. Huitema, and M.B. Sevryuk, *Quasi–periodic motions in families of dynamical systems*, Lecture Notes in Matematics, vol. 1645, Springer, Heidelberg, 1996.

[9] H.W. Broer, G.B. Huitema, F. Takens, and B.L.J. Braaksma, *Unfoldings and bifurcations of quasi–periodic tori*, Memoirs of the AMS, vol. 421, AMS, 1990.

[10] H.W. Broer, R. Roussarie, and C. Simò, *On the Bogdanov-Takens bifurcation for planar diffeomorphisms*, Proceedings of Equadiff '91 (C. Perello, C. Simò, and J. Sola-Morales, eds.), Springer Verlag, 1993, pp. 81–92.

[11] ———, *Invariant circles in the bogdanov-takens bifurcation for diffeomorphisms*, Ergodic Theory and Dynamical Systems **16** (1996), 1147–1172.

[12] T. Carletti and S. Marmi, *Linearization of analytic and non-analytic germs of diffeomorphisms of* $(\mathbb{C}, 0)$., Bulletin de la Société Mathematique de France **128** (2000), no. 1, 69–85.

[13] A. Chenciner, *Bifurcations de points fixes elliptiques. I. Courbes invariantes*, Publications Mathématiques de l'I.H.E.S. **61** (1985), 67–127.

[14] M.C. Ciocci, *Bifurcations of periodic points in families of reversible mappings and of quasi-periodic solutions in reversible systems*, Ph.D. thesis, Rijksuniversiteit Gent, 2003.

[15] G. Gentile and M. Procesi. *Periodic solutions for a class of nonlinear partial differential equations in higher dimension*, Preprint, arXiv:0804.4365v1, 2008.

[16] A. González-Enríquez, A. Haro and R. de la Llave. *Translated torus theory with parameters and applications*, In preparation.

[17] H.H. de Jong, *Quasi–periodic breathers in systems of weakly coupled pendulums*, Ph.D. thesis, Rijksuniversiteit Groningen, 1999.

[18] H. Hanßmann, *The quasi-periodic centre-saddle bifurcation*, Journal of Differential Equations **142** (1998), 305–370.

[19] ———, *Local and Semi-Local Bifurcations in Hamiltonian Dynamical Systems*, Lecture Notes in Mathematics, vol. 1893, Springer, 2007.

[20] J. Hoo, *Quasi-periodic bifurcations in a strong resonance: combination tones in gyroscopic stabilisation*, Ph.D. thesis, Rijksuniversiteit Groningen, January 2005.

[21] L. Hörmander, *An introduction to complex analysis in several variables*, Elsevier, Amsterdam, 1990, Third edition.

[22] ———, *The analysis of Linear Partial Differential Operators I. Distribution Theory and Fourier Analysis* , Springer, Heidelberg, 1990, Second edition.

[23] J. Moser, *Convergent series expansions for quasi-periodic motions*, Mathematische Annalen **169** (1967), 136–176.

[24] V. Poenaru, *Analyse différentielle*, LNM, vol. 371, Springer, Heidelberg, 1974.

[25] G. Popov, *Invariant tori, effective stability, and quasimodes with exponentially small error terms. I: Birkhoff normal forms.*, Annales de l'Institut Henri Poincaré **1** (2000), 223–248.

[26] ———, *Invariant tori, effective stability, and quasimodes with exponentially small error terms. II: Quantum Birkhoff normal forms.*, Annales de l'Institut Henri Poincaré **1** (2000), 249–279.

[27] ———, *KAM theorem for Gevrey Hamiltonians*, Ergodic Theory and Dynamical Systems **24** (2004), no. 5, 1753–1786.

[28] J. Pöschel, *Integrability of Hamiltonian systems on Cantor sets*, Communications of Pure and Applied Mathematics **35** (1982), 653–696.

[29] A.S. Pyartli, *Diophantine approximations on submanifolds of Euclidean space*, Functional Analysis and Applications **3** (1969), 303–306.

[30] R. Roussarie and F.O.O. Wagener, *A study of the Bogdanov-Takens bifurcation*, Resenhas IME-USP **2** (1995), 1–25.

[31] H. Rüssmann, *Kleine Nenner I: Über invariante Kurven differenzierbarer Abbildungen eines Kreisringes*, Nachr. Ak. Wiss. Gött., II Math. Phys. Kl. **5** (1970), 67–105.

[32] ———, *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 (J. Moser, ed.), LNP, no. 38, Springer, Heidelberg, 1975, pp. 598–624.

[33] M.B. Sevryuk, *KAM theory: persistence and smoothness*, Nonlinearity **21** (2008), T177–T185.

[34] E.M. Stein, *Singular integrals and the differentiability properties of functions*, Princeton, 1970.

[35] F. Takens, *Forced oscillations and bifurcations*, Communications of the Mathematical Institute 2, Rijksuniversiteit Utrecht, 1974.

[36] F.O.O. Wagener, *A note on Gevrey regular KAM theory and the inverse approximation lemma*, Dynamical Systems **18** (2003), 159–163.

[37] ———, *On the quasi-periodic d-fold degenerate bifurcation*, Journal of Differential Equations **216** (2005), 261–281.

[38] J. Xu and J. You, *Gevrey-smoothness of invariant tori for analytic nearly integrable hamiltonian systems under rüssmann's non-degeneracy condition*, Journal of Differential Equations **235** (2007), 609–622.

[39] E. Zehnder, *Generalized implicit function theorems with applications to some small divisor problems I*, Communications in Pure and Applied Mathematics **28** (1975), 91–140.

[40] D. Zhang and J. Xu, *Gevrey-smoothness of elliptic lower-dimensional invariant tori in Hamiltonian systems under Rüssmann's non-degeneracy condition*, Journal of Mathematical Analysis and Applications **323** (2006), no. 1, 293–312.

[41] ———, *On elliptic lower dimensional tori for Gevrey-smooth Hamiltonian systems under Rüssmann's non-degeneracy condition*, Discrete and Continuous Dynamical Systems **16** (2006), no. 3, 635–655.

[42] ———, *Invariant tori for Gevrey-smooth Hamiltonian systems under Rüssmann's non-degeneracy condition*, Nonlinear Analysis **67** (2007), 2240–2257.

[43] ———, *Invariant hyperbolic tori for gevrey-smooth hamiltonian systems under russmann's non-degeneracy condition*, Acta Mathematica Sinica-English Series **24** (2008), no. 10, 1625–1636.

Received xxxx 20xx; revised xxxx 20xx.

E-mail address: `F.O.O.Wagener@uva.nl`