*Forum of Mathematics, Sigma* (2026), **Vol. 14**:e72 1–42
doi:10.1017/fms.2026.10207

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

# KAM via standard fixed point theorems

Thomas Alazard[1] and Chengyang Shao[2]

[1]Centre de Mathématiques Laurent Schwartz (CMLS), CNRS, École polytechnique, France;
E-mail: thomas.alazard@polytechnique.edu.
[2]Institut des Hautes Études Scientifiques (IHÉS), France; E-mail: shao@ihes.fr (Corresponding author).

**Received:** 28 December 2025; **Accepted:** 17 January 2026

**2020 Mathematics Subject Classification:** *Primary* – 35S50; *Secondary* – 70H08

**Abstract**

With a mere usage of well-established properties of paradifferential operators, the conjugacy equations in several model KAM problems are converted to *parahomological equations* solvable by standard fixed point argument. Such discovery greatly simplifies KAM proofs, renders the traditional KAM iteration steps unnecessary, and may suggest a systematic scheme of finding quasi-periodic solutions of realistic magnitude.

## Contents

**1 Introduction** **2**
1.1 Organization of the paper . . . 2
1.2 A brief review of history . . . 5
1.3 Notation and convention . . . 7
**2 Circular map as illustrative model** **8**
2.1 Description of the problem . . . 9
2.2 Approximate right inverse . . . 10
2.3 Quick introduction to paradifferential calculus . . . 11
2.4 Parahomological equation . . . 12
2.5 Refinement in regularity . . . 14
2.6 Heuristics about generality . . . 16
**3 Quantitative paraproduct and paralinearization estimates** **18**
3.1 Meyer multipliers . . . 18
3.2 Quantitative estimates for paraproduct operators . . . 20
3.3 Bony's paralinearization theorem . . . 22
**4 Existence of invariant torus** **25**
4.1 Algebraic set-up . . . 25
4.2 Deriving parahomological equation . . . 29
4.3 Proof of Theorem 1.1–1.2 . . . 32
4.4 Miscellaneous . . . 35
**A "Direct method" for conjugacy problems** **37**
**References** **39**

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

## 1. Introduction

### 1.1. Organization of the paper

*Can regularity loss caused by small denominators be overcome by a standard functional analysis argument, instead of a Newtonian/Nash-Moser scheme?*

This has been a long-standing question in the field of dynamical systems, widely assumed to have a negative outcome. Surprisingly, we provide an affirmative answer in this paper: for several model KAM problems, it is possible to reduce the conjugacy equation to standard fixed point form via paralinearization techniques developed by Bony [11]. As a result, the "KAM steps" commonly considered essential to the problems are, in fact, unnecessary.

Several consequences follow immediately. Firstly, our new approach entirely bypasses the need for accelerated convergence techniques, a common feature in previous KAM arguments, whether Newtonian or not, used to compensate for regularity loss. It leads to significant simplification of the proof for KAM-type results. Moreover, following standard fixed-point theorems, this approach should allow substantially larger perturbations compared to earlier methods, as the cost of accelerated convergence scheme is unacceptably small magnitude of perturbation. We anticipate that the paradifferential approach to KAM theory may shed light on a systematic scheme for finding quasi-periodic motions of *realistic, physical* magnitude. Furthermore, a fixed point form of the equation may also yield insights about *nonexistence*.

After a brief review of the history of this topic, we present our paradifferential approach to KAM theory according to the following arrangement.

Section 2 is an essentially self-contained explanation of the core idea of our approach. We use the problem of circular maps studied by Arnold [8] as an illustrative model. A Nash-Moser (modified Newton) type proof is briefly outlined. We then introduce the concept of *parahomological equation*, serving as the counterpart to the homological equation in conventional presentation of KAM theory. Formally being the paradifferential version of the latter, the regularity gain due to paradifferential calculus balances with the loss. This enables one to convert it to a fixed point equation, resembling the *parainverse equation* introduced by Hörmander [44]. We then provide a heuristic showing that such simplicifaction is not specific to one-dimensional problems, but rather pertains to very broad class of conjugacy problems.

In Section 3, we present all the paradifferential calculus tools necessary for our approach: boundedness of paraproduct operator, composition estimate of paraproducts, and a refined version of J.-M. Bony's paralinearization theorem. The operator norms and magnitude of remainders are all estimated in a quantitative manner.

Section 4 collects the main results of this paper. We use the quantitative paradifferential calculus estimates developed in Section 3 to give new proofs of Hamiltonian conjugacy theorems. Instead of implementing any Newtonian algorithm, we will directly write down the parahomological equation and solve it by standard fixed point techniques. This is in contrast to most of the existing literature.

In order to state these results, we first set up the basic geometric notations. We consider $\mathbb{R}^n$ as the set of *column* vectors. For simplicity, we identify the space of mappings from $\mathbb{T}^n$ to $\mathbb{R}^n$ with the space of tangent vector fields on $\mathbb{T}^n$. The variable $\theta \in \mathbb{T}^n$ will be used to denote the variable on the "abstract torus." The phase space will be $\mathbb{T}^n \times \mathbb{R}^n$, the variable of which is denoted as $(x, y) \in \mathbb{T}^n \times \mathbb{R}^n$. Vector fields on the phase space will also be considered as *column* vectors. We use $D$ to denote differentiation in $(x, y)$, $\nabla$ to denote gradient with respect to $(x, y)$ (giving rise to column vectors), and $\partial$ to denote differentiation in $\theta$ or $x$. If $u$ is a mapping from $\mathbb{T}^n$ to the phase space, then the differential $\partial u$ is understood as a matrix of $n$ columns and $2n$ rows:

$$\partial u = \begin{pmatrix} \partial u^x \\ \partial u^y \end{pmatrix} \in \boldsymbol{M}_{2n\times n}.$$

The phase space $\mathbb{T}^n \times \mathbb{R}^n$ is endowed with standard symplectic form $\mathrm{d}x \wedge dy$. The corresponding symplectic structure has matrix representation

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

$$J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix} \in \boldsymbol{M}_{2n\times 2n}.$$

Given a Hamiltonian function $h(x, y)$ on the phase space $\mathbb{T}^n \times \mathbb{R}^n$, we denote the Hamiltonian vector field corresponding to $h$ by

$$X_h := J\nabla h = \begin{pmatrix} \nabla_y h \\ -\nabla_x h \end{pmatrix}.$$

Define the "flat" embedding of the torus $\mathbb{T}^n$ to the phase space by

$$\zeta_0 : \mathbb{T}^n \ni \theta \mapsto \begin{pmatrix} \theta \\ 0 \end{pmatrix} \in \mathbb{T}^n \times \mathbb{R}^n. \tag{1.1}$$

The Hamiltonian function under consideration takes the form

$$h(x, y) := a_0(x) + \langle a_1(x), y\rangle + \frac{1}{2}\langle Q(x)y, y\rangle + O(|y|^3), \tag{1.2}$$

where $a_0$ is a scalar function, $a_1$ is a $\mathbb{R}^n$-valued function, and $Q$ is a symmetric matrix valued function. In other words, we consider Hamiltonian functions defined near the "flat" embedded torus.

We also fix a Diophantine frequency vector $\omega \in \mathbb{R}^n$: there is a $\tau > 0$ and $\gamma > 0$ such that

$$|k \cdot \omega| \geq \frac{\gamma}{|k|^\tau}, \quad \forall k \in \mathbb{Z}^n \setminus \{0\}. \tag{1.3}$$

Standard measure theoretic arguments ensure the abundance of Diophantine vectors in the sense of measure: if $\tau > n - 1$, then the set of those $\omega \in \mathbb{R}^n$ satisfying the Diophantine condition (1.3) with $\gamma$ exhausting all positive real numbers is a set of the first Baire category and of full Lebesgue measure. We then write

$$\nabla_\omega = \sum_{j=1}^{n} \omega_j \partial_j$$

for the differentiation along the parallel vector field $\omega$ on $\mathbb{T}^n$.

We now state the main theorems of the paper. The function spaces involved in these statements are classical continuously differentiable spaces $C^N$, the Zygmund spaces $C^r_*$, and the Sobolev spaces $H^s$. See Subsection 1.3 for the definition of these function spaces.

**Theorem 1.1** (Existence of Invariant Torus). *Fix $\tau > n - 1$. Let $s > 2\tau + 2 + n/2 + \varepsilon$ be a fixed index, set $r = s - n/2$, and set $N_{s+r}$ to be the least integer such that $N_{s+r} > s + r$. Let $\zeta_0$, $h$ and $\omega$ be as in (1.1)-(1.3), and suppose the mean value* $\operatorname{Avg} Q$ *is an invertible matrix. Set*

$$M_Q = \max\left(\left|(\operatorname{Avg} Q)^{-1}\right|, |Q|_{L^\infty}\right).$$

*Define "error of being invariant" and "error of being integrable" as*

$$e_0 := X_h(\zeta_0) - \nabla_\omega \zeta_0, \quad e_1 := Q - \operatorname{Avg} Q. \tag{1.4}$$

*There are constants $c_0, c_1, c_2$ depending on $|h|_{C^{N_{s+r}+3}}$ and $M_Q$ with the following property. If*

$$\left\|e_0\right\|_{H^{s+2\tau+\varepsilon}} \leq \gamma^4 c_0, \quad |e_1|_{C^r_*} \leq \gamma^2 c_1 \tag{1.5}$$

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

*then there is an embedding* $u : \mathbb{T}^n \mapsto \mathbb{T}^n \times \mathbb{R}^n$ *of class* $H^s$*, such that*

$$\|u - \zeta_0\|_{H^s} \leq c_2 \gamma^{-2} \|e_0\|_{H^{s+2\tau+\varepsilon}},$$

*and* $u(\mathbb{T}^n)$ *is an invariant torus for the flow of* $h$*:*

$$X_h(u) - \nabla_\omega u = 0 \tag{1.6}$$

We next state a translated existence theorem that allows variation of frequency.

**Theorem 1.2** (Translated Conjugacy). *Fix* $\tau > n - 1$*. Let* $s > 2\tau + 2 + n/2 + \varepsilon$ *be a fixed index, set* $r = s - n/2$*, and set* $N_{s+r}$ *to be the least integer such that* $N_{s+r} > s + r$*. Let* $\zeta_0, h, \omega$ *be as in Theorem 1.1. Without any nondegeneracy assumption for* $Q$*, define*

$$M_Q = |Q|_{L^\infty}.$$

*Write* $h_\xi(x, y) = h(x, y) + \xi \cdot y$ *for* $\xi \in \mathbb{R}^n$*. Still define "error of being invariant" and "error of being integrable" as*

$$e_0 := X_h(\zeta_0) - \nabla_\omega \zeta_0, \quad e_1 := Q - \operatorname{Avg} Q. \tag{1.7}$$

*There are constants* $c_0, c_1$ *depending on* $|h|_{C^{N_{s+r}+3}}$ *and* $M_Q$ *with the following property. If*

$$\left\|e_0\right\|_{H^{s+2\tau+\varepsilon}} \leq \gamma^4 c_0, \quad |e_1|_{C^r_*} \leq \gamma^2 c_1, \tag{1.8}$$

*then there is an embedding* $u : \mathbb{T}^n \mapsto \mathbb{T}^n \times \mathbb{R}^n$ *of class* $H^s$ *and a constant vector* $\xi \in \mathbb{R}^n$*, such that*

$$\|u - \zeta_0\|_{H^s} \leq c_2 \gamma^{-2} \|e_0\|_{H^{s+2\tau+\varepsilon}},$$

*and* $u(\mathbb{T}^n)$ *is an invariant torus for the flow of the modified Hamiltonian* $h_\xi$*:*

$$X_{h_\xi}(u) - \nabla_\omega u = 0 \tag{1.9}$$

**Remark 1.1.** It is not hard to generalize the results to other function spaces, for example Zygmund spaces $C^r_*$ or more general Besov spaces. Though only existence results are stated, we can actually prove uniqueness and continuous dependence on parameters in both cases under more restrictive smallness assumptions. See the end of Subsection 4.3. On the other hand, although the theorems are stated for almost integrable systems, the coverage in fact is much broader under standard symplectic geometric consideration. See the discussion in Subsection 4.4.

Let us explain the reason of choosing these particular KAM type theorems for validation of our fixed point approach. Compared to standard account of KAM theorems solving a symplectic diffeomorphism that brings the Hamiltonian function to a normal form, starting from an approximate solution is computationally more convenient in practical applications to realistic systems. See for example [20, 21], where the authors coducted *computer assisted proof* of KAM type results for restricted three body problems of Sun-Jupiter-asteroid with *realistic* physical parameters. It can be expected that a fixed point approach will yield improvements regarding these realistic physical systems, and may advance the estimate of threshold of validity.

In the early stages of KAM theory, Hamiltonian conjugacy theorems were commonly formulated as structural stability theorems of KAM normal forms. This necessitated suitable action-angle variables for the system to reduce it to a perturbative form, usually a challenging task. But if one directly searches for an invariant torus near a given "approximately invariant torus" in the phase space, such technical issue can be largely bypassed. This idea dates back was used in, for example, [19, 78, 77, 59].

While the coverage of Theorem 1.1–1.2 may appear narrower than the results in [59], it remains sufficiently broad to apply to a diverse range. Notably, Theorem 1.1 is the famous Kolmogorov invariant tori theorem (see [55, 6]) with strongest nondegeneracy condition. In fact, as shown in [10], once the

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

embedding $u$ is given, a symplectic coordinate close to the original one can be readily constructed near the isotropic torus $u(\mathbb{T}^n)$, under which $h$ assumes a normal form and $u(\mathbb{T}^n)$ is flattened to $\mathbb{T}^n \times \{0\}$. Similarly, Theorem 1.2 is in fact a special case of the "théorème de conjugaison tordue" (see Féjoz [29]), which further implies the iso-energetic KAM theorem as a corollary. The more general "théorème de conjugaison tordue" by Herman was proved in [28], which is a properly degenerate KAM theorem. It appears that the method of our paper still applies with only minor alteration. For discussion of coverage of Theorem 1.1–1.2 besides these obvious specific cases, see Subsection 4.4.

We briefly discuss the possibility of further generalizations. Theorem 1.2, with suitable adaptation and extension, may find compatibility with "KAM for PDEs," including the Craig-Wayne-Bourgain approach to Melnikov persistency theorem. See for example [57, 23, 15, 16, 17] (and [10] for the applicability of this formalism to infinite dimensions). Such an alignment could potentially introduce a novel approach of finding "large magnitude" KAM solutions to Hamiltonian PDEs. To the author's knowledge, the first paradifferential construction of *periodic* solutions of PDEs involving small denominators is carried out by Delort [24]. Generalization to quasi-periodic solutions is possible. Due to the technicality involved in these extensions, we confine ourselves proving Theorem 1.1–1.2 in this paper to keep the narration as simple as possible. Extensions to "KAM for PDEs" are discussed with full detail in the forthcoming work [3].

### *1.2. A brief review of history*

Given the extensive volume of literature associated to KAM theory and Nash-Moser techniques, it does not seem practical to provide a panoramic review within such limited space. We refer the reader to [4, 12, 22, 28, 36, 60, 73, 83] for comprehensive description of these topics. Here, we confine ourselves to a brief overview of those works directly related to the central issue of this paper; that is, whether KAM iteration can be substituded by standard fixed point argument.

KAM theory is generally recognized to originate from Kolmogorov's 1954 report [55], although the problems it addressed had already interested mathematicians and physicists as early as the time of Poincaré: *does the Poincaré-Lindstedt series converge in presence of small denominators?* Or equivalently, *do quasi-periodic solutions persist upon perturbation?* Kolmogorov suggested a positive answer to this question by introducing a sequence of canonical transformations constructed via a modified Newtonian scheme. Arnold gave a detailed yet technically different proof of Kolmogorov's theorem in [6], and extended its coverage in [7]. Both Kolmogorov and Arnold worked in space of analytic functions. In the meantime, Moser [67] was able to replace analyticity by finite differentiability in a parallel context. Hence comes the abbreviation *KAM theory*.

Realizing similarity between the Newtonian scheme used by Kolmogorov-Arnold and that used by Nash [71] to address the isometric embedding problem, Moser [68, 69] then extracted this set of methods and applied it to broader classes of nonlinear problems. Related works that set stage for a general statement in graded spaces include Sergeraert [79] and Hörmander [42, 43]. In the original context of dynamical conjugacy problems, Zehnder [85, 86, 87] fitted them into adapted "hard implicit function theorems." These works are commonly recognized as the origin of the *Nash-Moser implicit function theorem*, which soon found its strength in nonlinear analysis. See [36] for a systematic narration.

Nevertheless, from a practical point of view, if the goal is to systematically find periodic or quasi-periodic motions of realistic magnitude, then KAM/Nash-Moser type methods, based on Newtonian iteration, are far from satisfactory. In fact, even when solving for the zero of a single-variable function, the Newtonian method is well-known to be much more sensitive to initial estimate than ordinary fixed point schemes. The loss of regularity caused by small denominators for conjugacy problems significantly worsens the scenario. As Hénon observed [38], the original proof by Arnold yields an infamous $10^{-300}$ upper bound for strength of perturbation, which is physically not acceptable (see also [58]):

> Ainsi, ces théorèmes, bien que d'un très grand intérêt théorique, ne semblent pas pouvoir en leur état actuel être appliqués à des problèmes pratiques, où les perturbations sont toujours beaucoup plus grandes. . .

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

> Thus, these theorems, although of great theoretical interest, do not seem to be able to be applied to practical problems in their current state, where the perturbations are always much larger. . .

However, Hénon also pointed out that numerical evidence suggests the persistence of quasi-periodic solutions even in the presence of strong perturbations. Numerical evidence, though it never serves as rigorous proof, could indicate the potential extension of KAM proofs to encompass stronger perturbations. There have been some successful research efforts aimed at quantifying the validity of KAM proofs. Notable among them are [19, 20, 21], where the authors proved KAM stability for certain Hamiltonian systems, including restricted three body problems, with realistic physical parameter. The proof is *quantitative* and *computer assisted* to validate the Newtonian algorithm.

Apart from practical considerations, mathematicians are interested in avoiding Nash-Moser type schemes because they typically offer less insight into the nonlinear structure. For example, the original proof of local existence of the Ricci flow [37] or the mean curvature flow [33] employed the Nash-Moser method, given the highly degenerate parabolic operators arising from linearization. However, the DeTurck technique introduced in [25] elegantly resolved this degeneracy by fixing geometric gauge. Another example is the quasi-linear perturbation of wave equations. Klainerman [52] [53] proved global well-posedness for quasi-linear perturbation of wave equations in spatial dimension higher than 6. His method was a Nash-Moser type iteration involving smoothing operators that both truncates Fourier modes *and* long-time ranges to deal with loss of decay. But Klainerman and Ponce [51] soon discovered the loss of decay could be compensated by working in Banach spaces of appropriately decaying weights. Later Klainerman [54] introduced the *vector field method*, drastically simplifying the proof of global well-posedness for wave equations. More recent development is also worth mentioning: Hintz-Vasy [41] proved the nonlinear stability of Kerr-de Sitter spacetime by a Nash-Moser inverse function theorem, which was later found by Fang [27] to be replaceable by suitable bootstrap argument.

As for the isometric embedding problem itself, 30 years after Nash's original work, Günther [35] converted the isometric embedding problem into a standard implicit function form through careful manipulation with Laplacian. Thus the Nash-Moser method is *not* necessary for its original intended context. Almost simultaneously, Hörmander [44] recognized the connection between Nash-Moser methods and paradifferential calculus: they both involve dyadic decomposition of nonlinearity. Hörmander introduced *parainverse operators* for inverse function problems with loss of regularity. Although the alternative proof for Nash embedding theorem in [44] implies slightly weaker result than [35], the general method of the former not only addresses to the vague observation *Nash-Moser technique can usually be replaced by elementary methods*, but also has the potential of applying to other nonlinear problems.

There have also been alternative proofs of KAM type results. Eliasson's paper [26] revisited the Poincaré–Lindstedt series, directly proving its convergence by implementing the delicate cancellations among the coefficients. Other proofs of KAM theorem include [18], using renormalization group acting on frequency space; [49], using a multidimensional continued fractions algorithm; [75, 74], which replaced the Newtonian iteration scheme with a "slowly converging" one; and [13, 14] by rational approximation of the frequency vector. Being insightful from different aspects, these alternative proofs are all still based on various types of iteration with improved convergence.

In the recent decade, the idea of replacing modified Newtonian (Nash-Moser or KAM) iteration with paradifferential calculus regained some attention. To mention a few, there was a proof of local well-posedness for gravity-capillary water waves using Nash-Moser theorem in [66]. By paralinearizing the system, however, Alazard-Burq-Zuily [1] proved the local well-posedness under significantly lower regime of regularity. In the construction of periodic solutions for dispersive differential equations on the torus, the traditional approach often employs Nash-Moser type techniques. But Delort [24] successfully substituted these techniques with ordinary iterations, aided by paradifferential calculus. The idea also echoed in recent advances for the study of *Landau damping*, the parallel of KAM theory in statistical mechanics. Originally proved to exist by a Newtonian scheme [70], the result was then proved in [9] using paraproducts in Gevrey spaces instead.

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

Regarding the necessity of such replacement, we particularly highlight the work of Herman [40], where the conjugacy equation for circular diffeomorphism was elegantly transformed into a fixed-point problem using Schwarz derivatives. Being surprising enough, even more is true. The technique was significantly extended by Marmi, Moussa, and Yoccoz [61, 62] in their proof of finite codimensional structural stability for interval exchanging maps. Traditional KAM iteration faces limitations in such settings, since the obstructions of solution of the linear homological equation become more severe in function spaces of higher regularity, as demonstrated by Forni [30, 31]. However, reformulating the conjugacy equation as a fixed-point problem effectively bypasses this issue, a key insight that supports the arguments in [61, 62].

On the other hand, the Schwarz derivative technique appears to be highly specific to one-dimensional problems. The analogous problem on stability of invariant surfaces of geodesic flows on translation surfaces, proposed by Forni in [30], remained open until the very recent work of Forni himself [32]. Forni applied the method developed in this article to establish finite codimensional structural stability of geodesic flows on translation surfaces – a setting where traditional KAM iteration fails due to regularity constraints[1]. The crux of Forni's proof lies in transforming the conjugacy equation into a fixed-point form via paralinearization. By resolving the open problem nearly three decades after its original introduction, such achievement might convince the reader of the merit of the paradifferential approach introduced in this paper. It is the aim of this paper to rekindle the idea of "Nash-Moser/KAM replaced by paradifferential" and develop a KAM theory from this new perspective.

### *1.3. Notation and convention*

The notions in this subsection are standard and can be found in any textbook on harmonic analysis, for example [81]. Throughout the paper, we do not distinguish functions defined on $\mathbb{T}^n$ with functions defined on $\mathbb{R}^n$ that are $2\pi$-periodic with respect to each variable. We represent a distribution $u$ on $\mathbb{T}^n$ as a Fourier series:

$$u = \sum_{k\in\mathbb{Z}^n} \hat{u}(k) e^{ik\cdot x},$$

where the Fourier coefficient

$$\hat{u}(k) = \int_{\mathbb{T}^n} u(x) e^{-ik\cdot x}\, \mathrm{d}x.$$

We denote $\operatorname{Avg} u = \hat{u}(0)$ for the mean value of $u$ if $u \in L^1$. If $u \in L^2$, then the series converges in $L^2$.

We introduce the *Littlewood-Paley decomposition* as follows:

**Definition 1.1.** Fix a function $\varphi \in C_0^\infty(\mathbb{R}^n)$, with support in an annulus $\{1/2 \le |\xi| \le 2\}$, so that

$$\sum_{j=1}^{\infty} \varphi(2^{-j}\xi) = 1 - \psi(\xi), \quad \operatorname{supp}\psi \subset \{|\xi| \le 1\}.$$

The Littlewood-Paley decomposition of a distribution $u$ on $\mathbb{T}^n$ is the defined as

$$u = \Delta_0 u + \sum_{j\ge 1} \Delta_j u, \quad \text{where} \quad \Delta_j u = \sum_{k\in\mathbb{Z}^n} \varphi(2^{-j}k)\hat{u}(k) e^{ik\cdot x}, \quad j \ge 1, \tag{1.10}$$

while one fixes $\Delta_0 u = \hat{u}(0) = \operatorname{Avg} u$.

[1]The authors are grateful to Giovanni Forni for bringing this to their attention.

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

The partial sum operator $S_j$ is defined as

$$S_j = \sum_{l \leq j} \Delta_l, \quad j \geq 0,$$

while for $j \leq 0$ one just fixes $S_j = \Delta_0$.

The summand $\Delta_j u$ in Definition 1.1 is called *j'th building block*. It has Fourier support contained in the dyadic annulus $\{0.5 \cdot 2^j \leq |\xi| \leq 2 \cdot 2^j\}$. The speed of convergence of $S_j u$ to $u$ reflects the regularity of $u$ – and it is the content of *Littlewood-Paley theory* to study this connection.

For an index $s \in \mathbb{R}$, the *Sobolev space* $H^s(\mathbb{T}^n)$ consists of those distributions $u$ on $\mathbb{T}^n$ such that

$$\|u\|_{H^s} := \left( \sum_{k \in \mathbb{Z}^n} \left(1 + |k|^2\right)^s \left|\hat{u}(k)\right|^2 \right)^{1/2} < +\infty.$$

The space $(H^s, \|\cdot\|_{H^s})$ is a Hilbert space. If $s \in \mathbb{N}$, then

$$H^s(\mathbb{T}^n) = \left\{ u \in L^2(\mathbb{T}^n) : \forall \alpha \in \mathbb{N}^n,\ |\alpha| \leqslant s,\ \partial_x^\alpha u \in L^2(\mathbb{T}^n) \right\},$$

where $\partial_x^\alpha$ is the derivative in the sense of distribution of $u$. Moreover, when $s$ is not an integer, the Sobolev spaces coincide with those obtained by interpolation. If we define

$$\|u\|_s^2 := \sum_{j=0}^{\infty} 2^{2js} \left\|\Delta_j u\right\|_{L^2}^2,$$

then it follows from Plancherel theorem that $\|\cdot\|_{H^s}$ and $\|\cdot\|_s$ are equivalent.

We will be using the *Zygmund spaces* (also known as *Lipschitz spaces* in the literature) throughout the paper. For an index $r \in \mathbb{R}$, the Zygmund space $C_*^r$ consists of those distributions $u$ on $\mathbb{T}^n$ such that

$$|u|_{C_*^r} := \sup_{j \geq 0} 2^{jr} |\Delta_j u|_{L^\infty} < +\infty. \tag{1.11}$$

Direct manipulation with series implies $H^s \subset C_*^{s-n/2}$ for any $s \in \mathbb{R}$. For noninteger $r > 0$, the $C_*^r$-norm is equivalent to the Hölder norm with index $r$, that is, the norm

$$|u|_{C^{[r]}} + \sum_{|\alpha| = [r]} \sup_{x,y \in \mathbb{T}^n} \frac{\left|\partial^\alpha u(x) - \partial^\alpha u(y)\right|}{|x - y|^{r - [r]}}.$$

Here $[r]$ is the integer part of $r$. However, when $r$ is a natural number, the space $C_*^r$ is strictly larger than the classical space of Lipschitz continuous functions.

## 2. Circular map as illustrative model

In this section, we will use the circular map model studied by Arnold [8] to illustrate the core idea of our paper. Although a quite complete global theory under sharp regularity and number-theoretic assumptions is available (see for example [39, 84, 50, 47]), we find it illuminating to revisit the perturbation problem for this model from the very beginning.

Throughout this section, we identify the circle $\mathbb{S}^1 = \mathbb{R}/2\pi\mathbb{Z}$, and write $\varrho_\alpha : x \mapsto x + \alpha$ for the rotation of angle $\alpha$, where $\alpha \in (0, 2\pi)$. Addition of arguments will all be understood as modulo $2\pi$. To avoid unnecessary confusion in notation, from now on we shall use $\eta^\iota$ to denote the inversion of a nonlinear mapping $\eta$ (should it exist), and use $a^{-1}$ to denote the reciprocal of a number $a$.

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

### *2.1. Description of the problem*

Let $\alpha \in (0, 2\pi)$ be noncommensurable with $\pi$. Consider the following classical problem:

- Suppose $f : \mathbb{S}^1 \mapsto \mathbb{S}^1$ is smooth and is close to 0. Is the diffeomorphism $x \mapsto x + \alpha + f(x)$ smoothly conjugate to the rotation $\varrho_\alpha$?

Of course, an integrablity condition must be posed for $f$. Recall the classical theorem due to Denjoy:

**Theorem 2.1.** *For a $C^3$ orientation-preserving diffeomorphism of $\mathbb{S}^1$ to itself, if its rotation number $\alpha$ is not commensurable with $\pi$, then it is topologically conjugate to the rotation $\varrho_\alpha$.*

If $x \mapsto x + \alpha + f(x)$ has rotation number different from $\alpha$, then by Denjoy's theorem, it definitely cannot conjugate to $\varrho_\alpha$. On the other hand, Denjoy's theorem has no implication on smoothness of the conjugation, even if the known diffeomorphism is smooth. Therefore the question is not answered by Denjoy's theorem. We modify it as follows:

- Suppose $f : \mathbb{S}^1 \mapsto \mathbb{S}^1$ is smooth and is close to 0, such that the diffeomorphism $x \mapsto x + \alpha + f(x)$ still has rotation number $\alpha$. Is it smoothly conjugate to the rotation $\varrho_\alpha$?

Let us write the hypothetical conjugation as $\eta(x) = x + u(x)$, and try to solve the superficially more general conjugacy equation

$$\eta(x + \alpha) = \eta(x) + \alpha + f \circ \eta(x) - \lambda$$

for $\eta$ and the auxiliary real parameter $\lambda$[2]. A simple rearrangement converts this equation to

$$\Delta_\alpha u = f \circ (\mathrm{Id} + u) - \lambda, \quad \text{where} \quad \Delta_\alpha u := u \circ \varrho_\alpha - u. \tag{2.1}$$

We may also compose the equation with $(\mathrm{Id} + u)^\iota$, and reformulate equation (2.1) as

$$\left[\Delta_\alpha u\right] \circ (\mathrm{Id} + u)^\iota = f - \lambda. \tag{2.2}$$

We will discuss the technical difference between (2.1) and (2.2) in Appendix A.

To clarify the obstacles of solving the conjugacy problem, we may linearize either (2.1) or (2.2) at $(u, \lambda) = (0, 0)$ along direction $(v, \mu)$. Neglecting $f$ as well, we obtain the linear *homological equation*

$$\Delta_\alpha v + \mu = h.$$

This linearized equation is solved via Fourier transform: normalizing $\hat{v}(0) = 0$, the unique solution is

$$v(x) = \sum_{k \neq 0} \frac{\hat{h}(k) e^{ikx}}{e^{ik\alpha} - 1}, \quad \mu = \mathrm{Avg}\, h,$$

Consequently, in order that the conjugacy problem is solvable even at the linear level, an additional number-theoretic condition must be posed for the number $\alpha$. We require that $\alpha/\pi$ is Diophantine of type $(\tau, \gamma)$: there are $\tau > 0$ and $\gamma > 0$ such that

$$\left| \frac{q\alpha}{\pi} - p \right| \geq \frac{\gamma}{q^\tau}, \quad \forall p, q \in \mathbb{Z} \setminus \{0\}. \tag{2.3}$$

Such numbers are abundant. Indeed, Liouville's inequality asserts that if an algebraic number $\alpha$ is of degree $D$, then the inequality $|q\alpha - p| \geq c/q^{D-1}$ must hold for some $c$. A measure-theoretic argument asserts that with $\tau > 1$ fixed, the set of $(\tau, \gamma)$ Diophantine numbers with $\gamma$ exhausting all positive real numbers form a set of full Lebesgue measure and of first Baire category.

[2]It is elementary that if $x \mapsto x + \alpha + f(x)$ has rotation number $\alpha$, and the conjugacy equation has a solution $(\eta, \lambda)$, then $\lambda$ is necessarily zero. Introducing $\lambda$ makes the presentation simpler. See for example Proposition III.4.1.1 of [39].

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

Obviously, if $\alpha$ satisfies the Diophantine condition (2.3), then $\Delta_\alpha^{-1}$ is a well-defined operator mapping functions of mean zero on $\mathbb{S}^1$ to functions of mean zero, satisfying

$$\|\Delta_\alpha^{-1} f\|_{H^s} \leq C\gamma^{-1}\|f\|_{H^{s+\tau}}, \quad \text{if } \operatorname{Avg} f = 0. \tag{2.4}$$

Here $C$ is an absolute constant independent of $\gamma$ or $s$. This enables one to solve the linearized equation $\Delta_\alpha v + \mu = h$, at least for very regular right-hand-side. But due to this loss of information on regularity, a usual iterative scheme to solve (2.2) will necessarily terminate after finitely many steps.

### 2.2. *Approximate right inverse*

Since the operator $\Delta_\alpha^{-1}$ causes a loss of regularity that cannot be compensated by any known elliptic technique, it is natural to employ modified Newtonian iterative scheme to solve the conjugacy problem (2.1). At a first glance it appears to be friendlier than (2.2), but in fact its linearized operator only admits an *approximate right inverse*. This brings more technicality for the application of Nash-Moser scheme. Let us describe the strategy of resolving it.

We set the unknwon $U = (u, \lambda)$, and define a mapping

$$\mathscr{F}(f, U) = \Delta_\alpha u - f \circ (\mathrm{Id} + u) + \lambda.$$

The linearization of $\mathscr{F}$ at $U = (u, \lambda)$ along $V = (v, \mu)$ is

$$D_U \mathscr{F}(f, U)V = \Delta_\alpha v - f' \circ (\mathrm{Id} + u)v + \mu.$$

In the literature of dynamical systems, given $U = (u, \lambda)$, the linearized equation

$$D_U \mathscr{F}(f, U)V = h \tag{2.5}$$

for the unknown $V = (v, \mu)$ with a given right-hand-side $h$ is referred to as the *homological equation*.

In general, one only expects to solve the homological equation (2.5) *approximately*, since the operator $\Delta_\alpha - f' \circ (\mathrm{Id} + u)$ is hard to invert. To explain the meaning of being "approximately solvable," we write down a simple yet crucial identity, which is a direct consequence of the definition of $\mathscr{F}$:

$$f' \circ (\mathrm{Id} + u) = \frac{\Delta_\alpha u'}{1 + u'} - \frac{[\mathscr{F}(f, U)]'}{1 + u'}. \tag{2.6}$$

So in fact

$$\begin{aligned} D_U \mathscr{F}(f, U)V &= \Delta_\alpha v - \frac{\Delta_\alpha u' \cdot v}{1 + u'} + \frac{[\mathscr{F}(f, U)]'}{1 + u'} v + \mu \\ &= (1 + u' \circ \varrho_\alpha)\Delta_\alpha \Big( \frac{v}{1 + u'} \Big) + \frac{[\mathscr{F}(f, U)]'}{1 + u'} v + \mu. \end{aligned} \tag{2.7}$$

Thus, defining

$$\Psi(U)h = \left( (1 + u')\Delta_\alpha^{-1} \left[ \frac{h - \mu(h)}{1 + u' \circ \varrho_\alpha} \right], \mu(h) \right), \quad \text{where } \mu(h) = \frac{\operatorname{Avg}\big((1 + u' \circ \varrho_\alpha)^{-1} h\big)}{\operatorname{Avg}\big((1 + u' \circ \varrho_\alpha)^{-1}\big)},$$

there holds

$$D_U \mathscr{F}(f, U)\Psi(U)h - h = [\mathscr{F}(f, U)]' \Delta_\alpha^{-1} \left( \frac{h}{1 + u' \circ \varrho_\alpha} - \mu(h) \right). \tag{2.8}$$

Equality (2.8) is the property defining "approximate solvability": *the linear operator $\Psi(U)$ is an exact right inverse of $D_U \mathscr{F}(f, U)$ at a precise solution $U$ of $\mathscr{F}(f, U) = 0$.*

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

The loss of regularity caused by $\Delta_\alpha^{-1}$ is usually identified as the primary difficulty for solving $\mathscr{F}(f,U)=0$. A modified Newtonian scheme was implemented by Kolmogorov [55], Arnold [6, 7, 8] and Moser [68, 69]. Being technically different, the general idea shared by them is to define a sequence $U_k$ by inductively solving a sequence of homological equations:

$$U_{k+1}=U_k-S_k\big[\Psi(U_k)\mathscr{F}(f,U_k)\big]. \tag{2.9}$$

Here the operator $S_k$ is either the restriction operator to a smaller domain in case all functions involved are analytic, or a smoothing operator in case all functions are of finite differentiability. The quadratic convergence property of Newtonian scheme cancels the large constants so produced and ensures the convergence of $U_k$ to a genuine solution. Proving this convergence is where the complexity accumulates.

The issue with approximate invertibility was first noticed by Zehnder [85, 86, 87], who introduced adapted hard implicit function theorems (Nash-Moser type theorems) to fit such conjugacy problems into a unified formalism. The proofs of these hard implicit function theorems still rely on modified Newtonian schemes, essentially being (2.9). For the statement and detailed proof of Nash-Moser theorem with either exactly or approximately invertible linearized operator, see [85, 86, 87, 36]. Though taking various forms on various graded spaces, all versions of Nash-Moser type theorems share the common features of *restoring regularity through smoothing operators* and *ensuring convergence through quadratic property of Newtonian algorithm*. While being universal and powerful theoretically, the practical disadvantage of this approach is that the allowed magnitude of perturbation is usually extremely small. A different approach is required if the aim is to find solutions of physically reasonable size.

### *2.3. Quick introduction to paradifferential calculus*

Surprisingly, if paradifferential calculus is used to attack on (2.1), we can drastically simplify the proof of existence of solution. For simplicity, we do not present a panorama of paradifferential calculus in this article; instead, we list the propositions that suffice for our exposition, along with brief descriptions of the heuristics behind them. So let us first introduce the *paraproduct operators*, the central objects of this paper.

**Definition 2.1.** Fix Littlewood-Paley decomposition as in Definition 1.1. Given a distribution $a=a(x)$ on $\mathbb{T}^n$, the paraproduct operator $T_a=\mathbf{Op}^{\mathrm{PM}}(a)$ associated to $a$ is defined as, regardless of the meaning of convergence,

$$T_a u=\mathbf{Op}^{\mathrm{PM}}(a)u:=\sum_{j\geq 0}S_{j-3}a\cdot\Delta_j u. \tag{2.10}$$

For $j\geq 1$, notice that the summand $S_{j-3}a\cdot\Delta_j u$ in (2.10) has its Fourier support contained in the annulus $\{0.25\cdot 2^j\leq|\xi|\leq 2.25\cdot 2^j\}$. A standard construction is the *paraproduct decomposition*:

$$au=T_a u+T_u a+R_{\mathrm{PM}}(a,u),\quad R_{\mathrm{PM}}(a,u)=\sum_{j,k:|j-k|<3}\Delta_j a\cdot\Delta_k u,$$

as long as the right-hand-side is well-defined in some suitable function space. Such decomposition separates apart *high-low*, *low-high* and *high-high* frequency interactions in a given multiplication $au$, and is therefore widely used in the study of harmonic analysis and dispersive partial differential equations. We refer the reader to Section 4.4 of [34] for a comprehensive description of the role it plays in harmonic analysis. [48, 46, 80] are typical examples of application of these constructions in dispersive PDEs.

The following results on paraproduct operators are directly cited from [11] (see also Chapters 8-10 of [45]). They are the three fundamental ingredients of paradifferential calculus.

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

**Proposition 2.1** (Continuity of paraproduct, rough version). *If $a \in L^\infty(\mathbb{T}^n)$, then $T_a$ is a bounded linear operator from $H^s$ to $H^s$ for all $s \in \mathbb{R}$, and in fact*

$$\|T_a\|_{\mathcal{L}(H^s,H^s)} \lesssim_s |a|_{L^\infty}.$$

**Proposition 2.2** (Composition of paraproducts, rough version). *If $a, b \in C^r_*$ with $r > 0$, then $T_aT_b - T_{ab}$ is a bounded linear operator from $H^s$ to $H^{s+r}$ for all $s \in \mathbb{R}$. Furthermore, $T_aT_b - T_{ab}$ is continuously bilinear in $a, b \in C^r_*$:*

$$\big\|T_aT_b - T_{ab}\big\|_{\mathcal{L}(H^s,H^{s+r})} \lesssim_{s,r} |a|_{C^r_*}|b|_{C^r_*}.$$

**Proposition 2.3** (Paralinearization, rough version). *Suppose $r = s - n/2 > 0$, $u \in H^s(\mathbb{T}^n;\mathbb{R}^L)$. Let $N_{s+r}$ be the least integer such that $N_{s+r} > s + r$. Suppose $F = F(x,z) \in C^{N_{s+r}+2}(\mathbb{T}^n \times \mathbb{R}^L)$. Then there holds the following paralinearization formula:*

$$F(x,u) - F(x,0) = T_{F'_z(x,u)}u + \mathcal{R}_{\mathrm{PL}}\big(F,u\big)\dot{u} \in H^s + H^{s+r},$$

*where $\mathcal{R}_{\mathrm{PL}}\big(F,u\big)$ is a bounded linear operator from $H^s \mapsto H^{s+r}$, so that*

$$\|\mathcal{R}_{\mathrm{PL}}(F,u)u\|_{H^{s+r}} \le |F|_{C^{N_{s+r}+2}}\big(1 + \|u\|_{H^s}\big)\|u\|_{H^s}.$$

*If furthermore $F \in C^{N_{s+r}+2+k}$, then $\mathcal{R}_{\mathrm{PL}}\big(F,u\big)$ has $C^k$ dependence on $u \in H^sH^{s+r}$.*

Precise statements of these results can be found in Section 3. Let us describe how Proposition 2.1–2.3 should be interpreted. The reason that Bony introduced paradifferential calculus in [11] is to understand microlocal regularity for solutions of nonlinear partial differential equations. For this purpose, it is necessary to try to separate out the *most irregular part* out of a given nonlinear expression $F(x,u)$, which is the content of the paralinearization theorem 2.3. The terminology *paralinearzation* refers to the observation that the most irregular part, $T_{F'_z(x,u)u}u$ out of $F(x,u)$, is exactly given by the paradifferential operator corresponding to the *linearization* of $F$.

Bony observed that $T_a$ exhibits advantages over the usual product operator: the operator $M_a : u \mapsto au$ is bounded on all Sobolev spaces $H^s$ only if $a \in C^\infty$, while the boundedness of $T_a$ on $H^s$ relies on the *minimal* regularity assumption that $a \in L^\infty$, as indicated by Proposition 2.1. On the other hand, $T_a$ shares the same *algebraic structure* as the usual multiplication $M_a$, up to smoothing remainders. While the mapping $a \mapsto M_a$ obviously is an algebra embedding from the function algebra $C^r_*$ to the operator algebra $\mathcal{L}(H^s,H^s)$ (with $s \le r$), Proposition 2.2 shows that the mapping $a \mapsto T_a$ is an algebra homomorphism from the function algebra $C^r_*$ to the operator algebra $\mathcal{L}(H^s,H^s)$ modulo smoothing operators for any $s \in \mathbb{R}$.

Therefore, Proposition 2.1–2.3 provides a procedure that enables one to study a nonlinear expression $F(x,u)$ *as if it were linear*: by paralinearization, one separates out the most irregular part, which obeys the same algebraic laws of the usual (linear) pseudo-differential operators.

### *2.4. Parahomological equation*

In this subsection, we illustrate how the conjugacy equation (2.1) can be solved in an extremely simple manner, provided that the paradifferential tools from Subsection 2.3 are admitted as standard, as in the field of harmonic analysis and dispersive partial differential equations. To the best of the authors' knowledge, the proof presented below stands as the shortest and simplest proof of KAM theorem for the circular map model available in literature.

*Proof.* Simple solution for (2.1). We notice that the general Dirichlet approximate theorem implies $\tau \ge 1$. Fix $s \ge \tau + 1.5 + \varepsilon$. Suppose *in a priori* $u \in H^s$, so that by Sobolev embedding, we have $u \in C^r_*$

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

with $r = s - 0.5 > 1$. Suppose $f$ is sufficiently smooth, for example $f \in C^{N_{s+r}+2}$, where $N_{s+r}$ is the least integer strictly greater than $s + r$.

By Proposition 2.3, we have the paralinearization formula for $\mathscr{F}(f, U)$:

$$\begin{aligned}\mathscr{F}(f,U) &= \Delta_\alpha u - f - T_{f'\circ(\mathrm{Id}+u)}u - \mathcal{R}_{\mathrm{PL}}(f,u)\cdot u + \lambda \\ &= \Delta_\alpha u - T_{\Delta_\alpha u'/(1+u')}u - f + T_{[\mathscr{F}(f,U)]'/(1+u')}u - \mathcal{R}_{\mathrm{PL}}(f,u)\cdot u + \lambda \\ &= T_{(1+u'\circ\varrho_\alpha)}\Delta_\alpha T_{1/(1+u')}u - f + T_{[\mathscr{F}(f,U)]'/(1+u')}u - \mathcal{R}_{\mathrm{PL}}(f,u)\cdot u + R_1(u) + \lambda.\end{aligned} \tag{2.11}$$

Here we used the key identity (2.6) again. The last equality in (2.11) is valid since

$$[T_{1/(1+u')}u]\circ\varrho_\alpha = T_{1/(1+u'\circ\varrho_\alpha)}(u\circ\varrho_\alpha).$$

Note that $\varrho_\alpha$ commutes with any Fourier multiplier, which is the reason that this equality holds. The remainder

$$\begin{aligned}R_1(u) &= \mathcal{R}_{\mathrm{CM}}(1+a, 1+b_1)u + \mathcal{R}_{\mathrm{CM}}(1+a, b_2)u \\ &= \mathcal{R}_{\mathrm{CM}}(a, b_1)u + \mathcal{R}_{\mathrm{CM}}(a, b_2)u,\end{aligned}$$

where the last step is because that $\mathcal{R}_{\mathrm{CM}}(\cdot,\cdot)$ is bilinear, and

$$a = u'\circ\varrho_\alpha, \quad b_1 = \frac{-u'\circ\varrho_\alpha}{1+u'\circ\varrho_\alpha}, \quad b_2 = -\frac{\Delta_\alpha u'}{(1+u')(1+u'\circ\varrho_\alpha)} \quad \text{are all in } H^{s-1} \subset C_*^{\tau+\varepsilon}.$$

is produced when replacing $T_{ab}$ by $T_a T_b$ in view of Proposition 2.2. Formally, we obtain (2.11) simply by replacing all products with paraproducts in (2.7).

We then consider the *parahomological equation* for the unknown $U = (u, \lambda) \in H^s \times \mathbb{R}$:

$$T_{(1+u'\circ\varrho_\alpha)}\Delta_\alpha T_{1/(1+u')}u = f + \mathcal{R}_{\mathrm{PL}}(f,u)\cdot u - R_1(u) - \lambda. \tag{2.12}$$

Equation (2.12) collects all terms but the one linear in $\mathscr{F}(f, U)$ in (2.11). It can be directly reduced to fixed point form, which we shall refer as the *parainverse equation* following Hörmander [44]:

$$u = T^{-1}_{1/(1+u')}\Delta_\alpha^{-1}T^{-1}_{(1+u'\circ\varrho_\alpha)}[f + \mathcal{R}_{\mathrm{PL}}(f,u)\cdot u - R_1(u) - \lambda]. \tag{2.13}$$

The equation itself uniquely determines $\lambda$: it should balance out the mean value to make $\Delta_\alpha^{-1}$ applicable.

By Proposition 2.2, the remainder

$$R_1(u) \in H^{s+r-1} \subset H^{s+\tau+\varepsilon},$$

is continuous in $u \in H^s$, and vanishes quadratically as $u \mapsto 0 \ni H^s$. By proposition 2.3,

$$\left\|\mathcal{R}_{\mathrm{PL}}(f,u)\cdot u\right\|_{H^{s+r}} \le C_s |f|_{C^{N_{s+r}+2}}\|u\|_{H^s}.$$

Due to Proposition 2.1 and (2.4), we find that the right-hand-side of (2.13) is a well-defined continuous mapping from $H^s$ to $H^{s+\varepsilon}$. The paradifferential remainder estimates further ensure that if $\|u\|_{H^s} \le \delta \ll 1$, then the $H^{s+\varepsilon}$ norm of the right-hand-side of (2.13) is controlled by

$$C\gamma^{-1}\left(|f|_{C^{N_{s+r}+2}}\delta + \delta^2\right).$$

Thus if we further require $\delta \ll \min(\gamma, 1)$ and $|f|_{C^{N_{s+r}+2}} \ll \delta$, the right-hand-side of (2.13) will be a continuous mapping from the closed ball $\bar{B}_\delta(0) \subset H^s$ to itself, with range actually in $H^{s+\varepsilon}$. By the Schauder fixed point theorem, such a mapping has a fixed point $u$. We thus have a solution $U = (u, \lambda)$ of the parainverse equation (2.13), hence the parahomological equation (2.12).

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

Given this solution $U$, the paralinearization formula (2.11) yields

$$\mathscr{F}(f,U) - T_{[\mathscr{F}(f,U)]'/(1+u')}u = 0.$$

If we consider the left-hand-side as a linear operator acting on $\mathscr{F}(f,U) \in C^1$, then by Proposition 2.1, it follows that

$$\big\|T_{[\mathscr{F}(f,U)]'/(1+u')}u\big\|_{H^s} \lesssim \big|\mathscr{F}(f,U)\big|_{C^1}\|u\|_{H^s}.$$

So if $|f|_{C^{N_{s+r}+2}}$ is sufficiently small (hence the solution $u$ has small $H^s$ norm), a Neumann series argument forces $\mathscr{F}(f,U) = 0$. □

**Remark 2.1.** Schauder's fixed point theorem implies the existence of solution in a nonconstructive manner. If we require more differentiability of $f$, for example $f \in C^{N_{s+r}+3}$ instead of $C^{N_{s+r}+2}$, then the remainder $\mathcal{R}_{\mathrm{PL}}(f,u)\cdot uu$ will be continuously differentiable in $u$. In fact, we can directly compute (see, e.g., [44]) the linearization of paralinearization remainder along an increment $v$ as

$$\begin{aligned} D_u[\mathcal{R}_{\mathrm{PL}}(F,u)\cdot u]v &= F'(u)v - T_{F''(u)v}u - T_{F'(u)}v \\ &= T_v(\mathcal{R}_{\mathrm{PL}}(F',u)\cdot u) + \text{smoother terms}. \end{aligned}$$

If the proof above, if $f \in C^{N_{s+r}+3}$, then $u \mapsto \mathcal{R}_{\mathrm{PL}}(f,u)\cdot u$ is a $C^1$ mapping from $H^s$ to $H^{s+r}$. Given that $f$ is close to 0 in $C^{N_{s+r}+3}$, a Banach fixed point argument is then applicable, since the right-hand-side of (2.13) will have Lipschitz constant $< 1$ in $u$. The solution $u$ is thus the limit of a Banach fixed point iteration sequence. Furthermore, if we assume $f \in C^\infty$, then by differentiating the fixed point equation (2.1), we obtain a *linear* iterative sequence that converges to the derivative of $u$. This easily yields the additional regularity $u \in C^\infty$.

Formally, (2.12) is the paradifferential counterpart of the linearized problem (2.5). Nevertheless, *the gain of regularity through paralinearization balances with the loss*, and thus renders unnecessary *any* convergence-improvement technique. This significantly differs from the prevailing practice in the existing KAM theory literature, with perhaps the only exception being Herman's Schwarz derivative technique; see the discussion in Subsection 1.2.

We note that the usual KAM/Nash-Moser iteration (2.9) relies on smoothing operators that are essentially Fourier truncations, which are also used in the construction of paradifferential operators as seen in (2.1). However, smoothing operators are not exploited to its maximal potential in KAM/Nash-Moser iteration since interactions of different frequencies are not used to balance the regularity loss caused by $\Delta_\alpha^{-1}$. This is exactly captured by paraproduct operators and paralinearization.

Furthermore, we emphasize that paraproduct operators, together with the remainders in paralinearization, all admit very explicit expressions, essentially involving only convolutions. Consequently, the size of the perturbation $f$ in (2.13), hence in the original conjugacy problem (2.1), does not have to be unreasonably small, as the proof of Banach or Schauder fixed point theorem indicates.

### 2.5. *Refinement in regularity*

If we take into account that $x+u(x)$ is a diffeomorphism of $\mathbb{S}^1$, then the loss of regularity can be managed more delicately using *paracomposition operator* introduced by Alinhac [5].

**Definition 2.2.** Let $\chi : \mathbb{T}^n \mapsto \mathbb{T}^n$ be a Lipschitz diffeomorphism. The paracomposition operator $\chi^\star$ associated to $\chi$ is defined by

$$\chi^\star F := \sum_{j\geq 0}(S_{j+N} - S_{j-N})\big((\Delta_j F)\circ\chi\big).$$

Here $N$ depends only on $|\partial\chi|_{L^\infty}$.

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

In fact, the paracomposition operator admits several equivalent definitions, see for example Appendix A of Chapter 2 in [82] or Section 5 of [2]; but we do not aim to elaborate them within our current paper. We directly cite several results concerning regularity of paracomposition operators. The proof can be found in, for example, [5], Appendix A of Chapter 2 in [82], Section 3 of [72], or Section 5 of [76]. Roughly speaking, the paracomposition operator captures the "irregularity" of the paralinearization remainder $F(u) - T_{F'(u)}u$.

**Proposition 2.4.** *If $\chi : \mathbb{T}^n \mapsto \mathbb{T}^n$ is a Lipschitz diffeomorphism, then $\chi^\star$ is a bounded linear operator from $H^s$ to $H^s$ for all $s \in \mathbb{R}$:*

$$\|\chi^\star\|_{\mathcal{L}(H^s,H^s)} \leq K_s\big(|\partial\chi|_{L^\infty}, |\partial\chi^{-1}|_{L^\infty}\big).$$

*Furthermore, given $F \in H^s(\mathbb{T}^n)$, the mapping $\chi \mapsto \chi^\star F$ is continuous from $W^{1,\infty}$ to $H^s$.*

**Proposition 2.5.** *Suppose $\rho > 1$, $\chi : \mathbb{T}^n \mapsto \mathbb{T}^n$ is a $C^\rho_*$ diffeomorphism, and $F \in H^{s+1}(\mathbb{S}^1)$ with $r := s - n/2 > 0$. Then the following paralinearization formula holds:*

$$F \circ \chi = \chi^\star F + T_{F'\circ\chi}\chi + \mathcal{R}_A(\chi)F,$$

*where*

$$\big\|\mathcal{R}_A(\chi)F\big\|_{H^{\rho+\min(\rho,r)}} \leq K_s\big(|\chi|_{C^\rho_*}, |\chi^{-1}|_{C^1}\big)\|F\|_{H^{s+1}}.$$

*Furthermore, given $F \in H^s(\mathbb{T}^n)$ and $\varepsilon > 0$, the mapping $\chi \mapsto \mathcal{R}_A(\chi)F$ is continuous from $C^\rho_*$ to $H^{s-\varepsilon}$.*

**Remark 2.2.** In fact, continuous dependence of paracomposition on the diffeomorphism $\chi$ was not explicitly proved in the literature cited above. However, one can directly read it off from the proof with a mere application of the dominated convergence theorem.

We can give a refined proof of sovability of (2.1). We still assume in a priori that $u \in H^s$, where $s \geq \tau + 1.5 + \varepsilon$, and $f \in H^{s+1}$. We then pick $r = \rho = s - 1/2$. By Proposition 2.5, with $\chi = \mathrm{Id} + u$, we have the paralinearization formula for $\mathscr{F}(f, U)$:

$$\begin{aligned}\mathscr{F}(f,U) &= \Delta_\alpha u - \chi^\star f - T_{f'\circ\chi}\chi - \mathcal{R}_A(\chi)f + \lambda\\ &= \Delta_\alpha u - T_{\Delta_\alpha u'/(1+u')}u - \chi^\star f + T_{[\mathscr{F}(f,U)]'/(1+u')}u - \mathcal{R}_A(\chi)f + \lambda\\ &= T_{(1+u'\circ\varrho_\alpha)}\Delta_\alpha T_{1/(1+u')}u - \chi^\star f + T_{[\mathscr{F}(f,U)]'/(1+u')}u - \mathcal{R}_A(\chi)f + R_1(u) + \lambda.\end{aligned} \tag{2.14}$$

Here the computations are almost the same as those in (2.11): the remainder $R_1(u) \in H^{s+r-1} \subset H^{s+\tau+\varepsilon}$ and vanishes bilinearly when $H^s \ni u \mapsto 0$, while the paracomposition remainder $\mathcal{R}_A(\chi)f \in H^{2s-1}$ by Proposition 2.5.

We then consider the *parahomological equation* for the unknown $U = (u, \lambda) \in H^s \times \mathbb{R}$:

$$T_{(1+u'\circ\varrho_\alpha)}\Delta_\alpha T_{1/(1+u')}u = \chi^\star f + \mathcal{R}_A(\chi)f - R_1(u) - \lambda. \tag{2.15}$$

Equation (2.15) is in fact just (2.12) in a delicate form. The parainverse equation (2.13) then takes an equivalent form

$$u = T^{-1}_{1/(1+u')}\Delta_\alpha^{-1}T^{-1}_{(1+u'\circ\varrho_\alpha)}\big(\chi^\star f + \mathcal{R}_A(\chi)f - R_1(u) - \lambda\big). \tag{2.16}$$

Here the $\lambda$ is still uniquely determined to balance out the mean value, making $\Delta_\alpha^{-1}$ applicable.

Due to Proposition 2.4 and 2.5, we find that the right-hand-side of (2.16) is a well-defined continuous mapping from $H^s$ to $H^{s+\varepsilon}$. The paradifferential remainder estimates further ensure that if $\|u\|_{H^s} \leq \delta$ and $|u'|_{L^\infty} \leq \rho$, then the $H^{s+\varepsilon}$ norm of the right-hand-side of (2.13) is controlled by

$$K(\delta,\rho)\|f\|_{H^{s+\tau+\varepsilon}} + C\delta^2,$$

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

where $K$ is an increasing function. Thus if $\|f\|_{H^{s+\tau+\varepsilon}}$ is sufficiently small, the right-hand-side of (2.16) will be a continuous mapping from the closed ball $\bar{B}(0,\delta) \subset H^s$ to itself, with compact range. By the Schauder fixed point theorem, such a mapping has a fixed point. The rest of the argument will then be identical with Subsection 2.4.

Sharing the same spirit with Subsection 2.4, the paracomposition approach just presented obviously requires less regularity for $f$ when $\tau \leq 2.5$, at the price of being technically more complicated. To keep our narration as simple as possible, we do not elaborate this approach within this paper. See Appendix A for more discussion.

### *2.6. Heuristics about generality*

It turns out that the aforementioned method is valuable not only in one-dimensional scenarios. In fact, it encompasses all the features of general conjugacy problems of interest in dynamical systems. Roughly speaking, this is the paradifferential calculus version of Zehnder's heuristics in Section 5 of [86]. We refer the reader to Zehnder's paper for his original argument, and present here the paradifferential version.

The formalism of conjugacy problems is as follows. Given an open set $\mathfrak{B}$ in a Banach space and an infinite dimensional Lie group $\mathfrak{G}$, suppose there is a differentiable group action ("dynamical system")

$$\Phi : \mathfrak{B} \times \mathfrak{G} \mapsto \mathfrak{B}.$$

Then $\Phi$ is subjected to the algebraic relation

$$\Phi(f, \mathrm{Id}) = f, \quad \Phi(f, g_2 \circ g_1) = \Phi\big(\Phi(f, g_2), g_1\big). \tag{2.17}$$

This is simply a result of definition of group homomorphism.

Now suppose $\mathfrak{N} \subset \mathfrak{B}$ is a closed linear submanifold, usually being the *normal forms* of the dynamical system. The conjugacy problem then reads: *is* $\mathfrak{N}$ *structurally stable in this dynamical system?* That is, if $f \in \mathfrak{B}$ is close to $\mathfrak{N}$, is it possible to find an element $g \in \mathfrak{G}$ close to the identity, so that $\Phi(f, g) \in \mathfrak{N}$?

To answer this question, introduce the unknwon $u = (\gamma, \mathfrak{n})$, where $\gamma$ belongs to the Lie algebra of $\mathfrak{G}$ and $\mathfrak{n} \in \mathfrak{N}$. Suppose $\mathcal{E}$ is a local chart[3] near the identity of the group $\mathfrak{G}$. Define a mapping

$$\mathscr{F}(f, u) = \Phi(f, \mathcal{E}(\gamma)) - \mathfrak{n}.$$

The conjugacy problem is then equivalent to finding zeroes of $\mathscr{F}$. Suppose that an approximate solution $u_0 = (0, \mathfrak{n}_0)$ is already found, that is $\mathscr{F}(f, u_0)$ is close to 0. We then aim to look for $u$ close to 0 such that $\mathscr{F}(f, u) = 0$

Introduce the map $\chi_\gamma$ on the Lie algebra of $\mathfrak{G}$ by $\mathcal{E}(\chi_\gamma(\gamma_1)) = \mathcal{E}(\gamma) \circ \mathcal{E}(\gamma_1)$. Then (2.17) becomes

$$\Phi\big(f, \chi_\gamma(\gamma_1)\big) = \mathscr{F}\big(\Phi(f, \gamma), \gamma_1\big).$$

Denote by $v = (\psi, \mathfrak{m})$ the increment of $u = (\gamma, \mathfrak{n})$, and set $L_\gamma v = (D\chi_\gamma(0)\psi, \mathfrak{m})$. Differentiating with respect to $u$, evaluating at $\gamma_1 = 0$, we obtain

$$D_u\mathscr{F}(f, u)L_\gamma v = D_u\mathscr{F}\big(\Phi(f, \gamma), (0, \mathfrak{n})\big)v.$$

Thus by Taylor's formula,

$$\begin{aligned} D_u\mathscr{F}(f, u)L_\gamma v &= D_u\mathscr{F}\big(\mathfrak{n}, (0, \mathfrak{n})\big)v + \big(D_u\mathscr{F}\big(\Phi(f, \gamma), (0, \mathfrak{n})\big) - D_u\mathscr{F}\big(\mathfrak{n}, (0, \mathfrak{n})\big)\big)v \\ &= D_u\mathscr{F}\big(\mathfrak{n}, (0, \mathfrak{n})\big)v + \boldsymbol{B}\big(\mathscr{F}(f, u), v\big), \end{aligned}$$

[3]This local chart usually is not the exponential map, since the exponential map is not surjective for an infinite dimensional Lie group. See for example, Section 42 of [56].

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

where $\boldsymbol{B}$ vanishes bilinearly near (0,0). Equivalently,

$$D_u\mathscr{F}(f,u)v = D_u\mathscr{F}\big(\mathfrak{n},(0,\mathfrak{n})\big)L^1_\gamma v + \boldsymbol{B}\big(\mathscr{F}(f,u),L^1_\gamma v\big). \tag{2.18}$$

The idea of paradifferential approach to the conjugacy equation $\mathscr{F}(f,u)=0$ is then based on the algebraic identity (2.18). Suppose the Banach space $\mathfrak{B}$ and Lie group $\mathfrak{G}$ are realized as functions and (finite-dimensional) diffeomorphisms, and the mapping $\mathscr{F}(f,u)$ is a classical nonlinear differential operator acting on $u$, in the sense that it is a function of $(u,\partial u,\partial^2 u,\cdots)$. Then the paralinearization theorem of Bony [11] should imply

$$\begin{aligned}\mathscr{F}(f,u) &= \mathscr{F}(f,u_0)+T_{D_u\mathscr{F}(f,u)}(u-u_0)+\boldsymbol{R}_1(u-u_0)\\ &= T_{D_u\mathscr{F}\left(\mathfrak{n},(0,\mathfrak{n})\right)}T_{L^1_\gamma}(u-u_0)+\mathscr{F}(f,u_0)+\boldsymbol{R}_2(u-u_0)+\boldsymbol{B}\big(T_{\mathscr{F}(f,u)},T_{L^1_\gamma}(u-u_0)\big).\end{aligned} \tag{2.19}$$

Here $u=(\gamma,\mathfrak{n})\simeq(0,\mathfrak{n}_0)$, and $\boldsymbol{R}_1,\boldsymbol{R}_2$ are smoother remainders produced by paralinearization. The somewhat abused notation $\boldsymbol{B}\big(T_{\mathscr{F}(f,u)},T_{L^1_\gamma}(u-u_0)\big)$ stands for the replacement of "product" $\boldsymbol{B}\big(\mathscr{F}(f,u),L^1_\gamma(u-u_0)\big)$ by the corresponding "paraproduct" $\boldsymbol{B}\big(T_{\mathscr{F}(f,u)},T_{L^1_\gamma}(u-u_0)\big)$, should $\boldsymbol{B}$ involve only usual functional operations.

The following step is to make sure that the *parahomological equation*

$$T_{D_u\mathscr{F}\left(\mathfrak{n},(0,\mathfrak{n})\right)}T_{L^1_\gamma}(u-u_0)+\mathscr{F}(f,u_0)+\boldsymbol{R}_2(u-u_0)=0 \tag{2.20}$$

is solvable. We will actually consider its equivalent fixed point form, refered as the *parainverse equation*:

$$(u-u_0)=-T^{-1}_{L^1_\gamma}T^{-1}_{D_u\mathscr{F}\left(\mathfrak{n},(0,\mathfrak{n})\right)}\big(\mathscr{F}(f,u_0)+\boldsymbol{R}_2(u-u_0)\big),$$

and make sure that it is solvable.

For most conjugacy problems, this really is the case, *even if inverting* $D_u\mathscr{F}\big(\mathfrak{n},(0,\mathfrak{n})\big)$ *causes a loss of regularity due to small denominator*. In fact, that loss of regularity can be compensated by the additional smoothness of the paradifferential remainder $\boldsymbol{R}_2(u-u_0)$. Thus, if $\mathscr{F}(f,u_0)$ is sufficiently smooth, the right-hand-side of the fixed point equation defines a mapping with *no* loss of regularity (e.g., from $H^s$ to $H^s$). We note that this is the paradifferential version of Zehnder's "finding approximate right inverse." Consequently, assuming $\mathscr{F}(f,u_0)$ is close enough to 0, a standard fixed point argument should produce a solution $u\simeq u_0$ of (2.20).

Once a solution $u$ to the parahomological equation (2.20) is found, the first three terms in the right-hand-side of (2.19) are then cancelled, and the paralinearization formula then becomes

$$\mathscr{F}(f,u)=\boldsymbol{B}\big(T_{\mathscr{F}(f,u)},L^1_\gamma(u-u_0)\big)\approx T_{\mathscr{F}(f,u)}L^1_\gamma(u-u_0).$$

Now since a paradifferential operator $T_a$ depends only on very mild regularity of the symbol $a$, this equation must imply $\mathscr{F}(f,u)=0$ by a Neumann series argument, provided that $u$ is close to $u_0$.

*To conclude, solution of the parahomological equation in fact solves the original conjufacy problem.*

Of course, the steps outlined above are merely heuristics. In practice, each of the assumptions they rely upon must be validated in a case-by-case fashion. Nevertheless, given that these intuitive deductions pertain to very general scenarios, it is reasonable to expect that they adequately cover several well-studied conjugacy problems. We point out, omitting most details, that the following model problems all fit into this formalism: conjugation of standard map (see, e.g., Section 2 of [60]), structural stability of parallel vector field on torus (see, e.g., [74]), existence of invariant torus of symplectic map (see, e.g., [59]), and conjugation of Hamiltonian vector fields. We will elaborate on the last one in Section 4.

The idea of replacing a linearized equation by its paradifferential version dates back to Hörmander's paper [44], where Hörmander proposed the terminology *parainverse*. Hörmander only considered nonlinear mappings with *exactly* invertible linearized operators, instead of *approximately* invertible

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

ones. Nevertheless, it is still appropriate to attribute the core idea of the aforementioned heuristics to [44] (and [86]).

## 3. Quantitative paraproduct and paralinearization estimates

In this section, we present quantitative estimates concerning paraproducts and paralinearization. They are refinements of standard contents in modern harmonic analysis; see for example, [11, 64, 65], Chapter 9-10 of [45], or Chapter 3-4 of [63]. From a harmonic analysis point of view, paradifferential operators fall into special subclasses of pseudo-differential operators of "forbidden" type (1,1), but still enjoy a symbolic calculus as usual (1,0) pseudo-differential operators do. However, to keep prerequisites minimal, we attempt to give neither comprehensive exposition of general (1,1) pseudo-differential operators nor the symbolic calculus, but confine ourselves with what is necessary.

### *3.1. Meyer multipliers*

Before providing the quantitative paradifferential operator estimates, let us first state some auxiliary results relating to *Meyer multipliers*. Formally, given a sequence $\{m_j(x)\}_{j\geq 0}$ of functions on $\mathbb{T}^n$, we call the linear operator

$$f(x) \mapsto \sum_{j\geq 0} m_j(x)\cdot(\Delta_j f)(x)$$

the *Meyer multiplier* associated to $\{m_j\}_{j\geq 0}$. Here the $\Delta_j$'s are the Littlewood-Paley building block operators introduced in (1.10). The fundamental theorem for such a linear operator is the following:

**Proposition 3.1** (Meyer multiplier). *Let $s, r \in \mathbb{R}$, and suppose that $s + r > 0$. Let $N_{s+r}$ be the smallest integer such that $N_{s+r} > s + r$. Suppose $(m_j)_{j\geq 0}$ is a sequence in $L^\infty(\mathbb{T}^n)$, such that for multi-indices $\alpha$ with $|\alpha| \leq N_{s+r}$, there holds*

$$\left|\partial^\alpha m_j\right|_{L^\infty} \leq M_\alpha 2^{j(|\alpha|-r)}.$$

*Then the linear operator*

$$f \mapsto \sum_{j=0}^{\infty} m_j \Delta_j f$$

*is bounded from $H^s(\mathbb{T}^n)$ to $H^{s+r}(\mathbb{T}^n)$. The norm of this operator is estimated by*

$$C_s \sum_{\beta:|\beta|\leq N_{s+r}} M_\beta.$$

We need a lemma concerning Sobolev regularity of series to aid the proof of Proposition 3.1. It is of some independent interest per se.

**Lemma 3.1.** *Let $s > 0$, let $N_s$ be the smallest integer such that $N_s > s$. There is a constant $C_s$ with the following properties. If $\{f_k\}_{k\geq 0}$ is a sequence in $H^{N_s}(\mathbb{T}^n)$, such that for any multi-index $\alpha$ with $|\alpha| \leq N_s$, there holds*

$$\|\partial^\alpha f_k\|_{L^2} \leq 2^{k(|\alpha|-s)} c_k, \quad (c_k) \in \ell^2(\mathbb{N}),$$

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

*then* $\sum_k f_k \in H^s(\mathbb{R}^n)$ *and*

$$\left\| \sum_{k\geq 0} f_k \right\|_{H^s}^2 \leq C_s \sum_{k\geq 0} c_k^2.$$

*Proof.* If $j \leq k$, then by taking $\alpha = 0$, obviously

$$\|\Delta_j f_k\|_{L^2} \leq \|f_k\|_{L^2} \leq 2^{-ks} c_k.$$

If $j > k$, then by Bernstein's inequality,

$$\|\Delta_j f_k\|_{L^2} \leq C_s 2^{-N_s j} \sum_{|\alpha|=N_s} \|\partial^\alpha f_k\|_{L^2} \leq C_s 2^{-ks} 2^{N_s(k-j)} c_k.$$

Thus, for $f = \sum_k f_k$, we have

$$2^{js}\|\Delta_j f\|_{L^2} \leq C_s \sum_{k:k<j} 2^{(N_s-s)(k-j)} c_k + \sum_{k:k\geq j} 2^{(j-k)s} c_k. \tag{3.1}$$

Given that $s > 0$ and $N_s > s$, the right-hand-side is the convolution of the $\ell^1$ sequences $(2^{-(N_s-s)k})$, $(2^{-sk})$ with the $\ell^2$ sequence $(c_k)$. Therefore, the right-hand-side forms an $\ell^2$ sequence (with index $j$), with norm controlled by $C_s\|(c_k)\|_{\ell^2}$. Thus $f \in H^s$. □

*Proof of Proposition 3.1.* In fact, for $|\alpha| \leq N_{s+r}$, each summand $m_j \Delta_j f$ satisfies

$$\begin{aligned}\left\|\partial^\alpha (m_j \Delta_j f)\right\|_{L^2} &\leq C_\alpha \sum_{\beta:\beta\leq\alpha} \left|\partial^\beta m_j\right|_{L^\infty} \left\|\partial^{\alpha-\beta}\Delta_j f\right\|_{L^2} \\ &\leq C_\alpha 2^{j(|\alpha|-r-s)} \left(\sum_{\beta:\beta\leq\alpha} M_\beta\right) 2^{js}\|\Delta_j f\|_{L^2}.\end{aligned}$$

We can then directly apply Lemma 3.1 to conclude. □

**Corollary 3.1.1.** *If the sequence* $\{f_k\}$ *in Lemma 3.1 in addition satisfies the spectral condition*

$$\operatorname{supp} \hat{f}_k \subset \{|\xi| \geq \delta 2^k\}$$

*for some* $\gamma > 0$*, then the restriction* $s > 0$ *in Lemma 3.1 can be relaxed to any real index s:*

$$\left\| \sum_{k\geq 0} f_k \right\|_{H^s}^2 \leq C_{s,\delta} \sum_{k\geq 0} c_k^2.$$

*Proof.* In fact, if the $f_k$'s satisfy the spectral condition, then the second sum in (3.1) becomes a finite one, controlled by $O(\log(\delta^{-1}))$. Thus the sequence $\sum_{k:k\geq j} 2^{(j-k)s} c_k$ automatically forms an $\ell^2$ sequence (with index $j$). □

**Remark 3.1.** Suppose the multipliers $m_j = m_j(\lambda)$ in Proposition 3.1 depend on a parameter $\lambda$ varying in some Banach space, such that

$$\max_{\alpha:|\alpha|\leq N_{s+r}} \sup_j 2^{-j(|\alpha|-r)} \left|\partial^\alpha D_\lambda m_j(\lambda)\right|_{L^\infty} < \infty.$$

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

Then we can simply repeat the proof of Proposition 3.1 to conclude the following: the operator-valued function

$$\lambda \mapsto \sum_{j=0}^{\infty} m_j(\lambda)\Delta_j \in \mathcal{L}(H^s, H^{s+r})$$

is differentiable in $\lambda$ under the norm topology of $\mathcal{L}(H^s, H^{s+r})$, whose differential is simply

$$\sum_{j=0}^{\infty} D_\lambda m_j(\lambda)\Delta_j.$$

### *3.2. Quantitative estimates for paraproduct operators*

With the aid of Meyer multiplier estimates, we are now able to provide a quantitative proof of boundedness of paraproduct operators. We will first list several properties related to Littlewood-Paley characterization of the Zygmund space. The proof is quite elementary using the definition of Zygmund spaces, as in (1.11).

**Proposition 3.2.** *Suppose $r > 0$ and $f \in C^r_*$. Then for any multi-index $\alpha$, there holds*

$$\left|\partial^\alpha S_j f\right|_{L^\infty} \lesssim_{r,\alpha} \begin{cases} 2^{j(|\alpha|-r)_+}|f|_{C^r_*} & |\alpha| \neq r \\ j|f|_{C^r_*} & |\alpha| = r \end{cases}$$

*where $s_+ = \max(s, 0)$. Furthremore, there holds*

$$|f - S_j f| \lesssim_r 2^{-jr}|f|_{C^r_*}.$$

We are now at the place to state and prove the continuity of paraproduct operators, which appeared as Proposition 2.1.

**Proposition 3.3.** *For all $a \in L^\infty(\mathbb{T}^n)$ and all $s \in \mathbb{R}$, the paraproduct $T_a$ is a bounded linear operator from $H^s(\mathbb{T}^n)$ to $H^s(\mathbb{T}^n)$:*

$$\|T_a\|_{\mathcal{L}(H^s,H^s)} \leq C_s|a|_{L^\infty}.$$

*Proof.* In fact, in the defining equality (2.10), each summand has Fourier support in the annulus $\{0.25 \cdot 2^j \leq |\xi| \leq 2.25 \cdot 2^j\}$. We then simply notice that the Meyer multipliers $m_j = S_{j-3}a$ satisfy

$$\left|\partial^\alpha m_j\right|_{L^\infty} \leq C_\alpha 2^{j|\alpha|}|a|_{L^\infty}$$

by Proposition 3.2. Thus we can repeat the proof of Corollary 3.1.1 and conclude. □

The nontrivial fact about paraproduct operators is that they behave as if they were genuine multiplication operators: the multiplication $T_a T_b$ is the same as $T_{ab}$ modulo smoothing operators. Formally, this means that $a \mapsto T_a$ is an algebra homomorphism from the algebra of functions to the algebra of bounded operators (modulo smoothing ones). We provide the precise version of Proposition 2.2:

**Proposition 3.4.** *Fix $r > 0$, and suppose $a, b \in C^r_*$. Then*

$$\mathcal{R}_{\mathrm{CM}}(a, b) := T_a \circ T_b - T_{ab}$$

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

*is a bounded linear operator from $H^s(\mathbb{T}^n)$ to $H^{s+r}(\mathbb{T}^n)$ for all $s \in \mathbb{R}$, where CM is the abbreviation for composition de paramultiplication, satisfying the tame estimate*

$$\|\mathcal{R}_{\mathrm{CM}}(a,b)\|_{\mathcal{L}(H^s,H^{s+r})} \leq C_{s,r}\big(|a|_{L^\infty}|b|_{C^r_*} + |a|_{C^r_*}|b|_{L^\infty}\big).$$

*In other words, $\mathcal{R}_{\mathrm{CM}} : C^r_* \times C^r_* \mapsto \mathcal{L}(H^s, H^{s+r})$ is a bounded bilinear mapping.*

*Proof.* The proof here is exactly the same as Theorem 2.3 of [11]. Suppose $a, b \in C^r_*$ and $u \in H^s$ are given. We set $v = T_b u$, $v_q = S_{q-3}b \cdot \Delta_q u$. By Corollary 3.1.1 and the definition (1.11) of Zygmund space norm, the error

$$R_1 v := T_a v - \sum_p S_{p-5}a \cdot \Delta_p v = \sum_p (\Delta_{p-4}a + \Delta_{p-3}a)\Delta_p u$$

satisfies $\|R_1 v\|_{H^{s+r}} \leq C_{s,r}|a|_{C^r_*}|b|_{L^\infty}\|u\|_{H^s}$, since each summand $(\Delta_{p-4}a + \Delta_{p-3}a)\Delta_p v$ has Fourier support contained in the annulus $\{0.25 \cdot 2^p \leq |\xi| \leq 2.25 \cdot 2^p\}$. We then compute

$$\begin{aligned} T_a T_b u - \sum_q (S_{q-5}a)(S_{q-3}b)\Delta_q u &= T_a T_b u - \sum_q \sum_{p:p\leq q-5} \Delta_p a \cdot v_q \\ &= \sum_q \sum_{p:p\leq q-5} \Delta_p a \cdot \big(\Delta_q v - v_q\big) + R_1 T_b u \\ &= \sum_{p\geq -5} \Delta_p a \sum_{q:p+5\leq q\leq p+7} \big(\Delta_q v - v_q\big) + R_1 T_b u. \end{aligned} \tag{3.2}$$

The last step is because $\sum_q v_q = \sum_q \Delta_q v$ and $v_q$ has Fourier support contained in the annulus $\{0.25\cdot 2^q \leq |\xi| \leq 2.25 \cdot 2^q\}$. Using Corollary 3.1.1 once again, we estimate (3.2) as

$$\left\| T_a T_b u - \sum_q (S_{q-5}a)(S_{q-3}b)\Delta_q u \right\|_{H^{s+r}} \leq C_{s,r}|a|_{C^r_*}|b|_{L^\infty}\|u\|_{H^s}. \tag{3.3}$$

On the other hand, if we set

$$c_q = (S_{q-5}a)(S_{q-3}b),$$

then $c_q$ has Fourier support contained in the ball $|\xi| \leq 0.25 \cdot 2^q$, and

$$\begin{aligned} |ab - c_q|_{L^\infty} &\leq |\big((\mathrm{Id} - S_{q-5})a\big)b|_{L^\infty} + |(S_{q-5}a)\big((\mathrm{Id} - S_{q-5})b\big)|_{L^\infty} \\ &\leq \sum_{p>q-5} 2^{-pr}|a|_{C^r_*}|b|_{L^\infty} + |a|_{L^\infty} \sum_{p>q-5} 2^{-pr}|b|_{C^r_*} \\ &\leq C_r 2^{-q}\big(|a|_{C^r_*}|b|_{L^\infty} + |a|_{L^\infty}|b|_{C^r_*}\big). \end{aligned} \tag{3.4}$$

Noticing that

$$T_{ab}u - \sum_q c_q \Delta_q u = \sum_q \big(S_{q-3}(ab) - c_q\big) \cdot \Delta_q u = \sum_q \big(S_{q-3}(ab) - ab + ab - c_q\big) \cdot \Delta_q u,$$

using Corollary 3.1.1 and (3.4), we find

$$\left\| T_{ab}u - \sum_q c_q \cdot \Delta_q u \right\|_{H^{s+r}} \leq C_{s,r}\big(|a|_{C^r_*}|b|_{L^\infty} + |a|_{L^\infty}|b|_{C^r_*}\big)\|u\|_{H^s}. \tag{3.5}$$

Combining (3.3) and (3.5), the proof is complete. □

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

### 3.3. Bony's paralinearization theorem

In this subsection, we present a quantitative refinement of Bony's paralinearization theorem in [11], which asserts that given a nonlinear composition $F(u)$, the paraproduct $T_{F'(u)}u$ carries most of the irregularity, and

$$F(u) - T_{F'(u)}u$$

is more regular than $F(u)$. This is a far-reaching generalization of the paraproduct decomposition.

While Bony's theorem is well-established and extensively applied in modern nonlinear analysis, we still need to manage the *size* of the smoothing remainder to effectively study the parahomological equation. We state the following refined version of Proposition 2.3:

**Proposition 3.5.** *Fix $s, r > 0$, and suppose $u \in (H^s \cap C^r_*)(\mathbb{T}^n; \mathbb{R}^L)$. Let $N_{s+r}$ be the smallest integer $> s + r$. Suppose $F = F(x, z) \in C^{N_{s+r}+2}(\mathbb{T}^n \times \mathbb{R}^L)$. Then there holds the following paralinearization formula:*

$$\begin{aligned} F(x,u) - F(x,0) &= T_{F'_z(x,u)}u + \mathcal{R}_{\mathrm{PL}}(F,u)\cdot u \\ &=: T_{F'_z(x,u)}u + \mathcal{R}_{\mathrm{PL};1}(F)\cdot u + \mathcal{R}_{\mathrm{PL};2}(F,u)\cdot u \in H^s + H^{s+r}. \end{aligned}$$

*Here $\mathcal{R}_{\mathrm{PL};1}(F) : H^s \mapsto H^{s+r}$ and $\mathcal{R}_{\mathrm{PL};2}(F,u) : H^s \mapsto H^{s+r}$ are bounded linear operators, so that for some increasing function $K_{r,s}$ depending only on $r, s$,*

$$\begin{aligned} \big\|\mathcal{R}_{\mathrm{PL};1}(F)\big\|_{\mathcal{L}(H^s,H^{s+r})} &\le C_{r,s}\big|F'_z(x,0) - \mathrm{Avg}\, F'_z(x,0)\big|_{C^r_*} \\ \big\|\mathcal{R}_{\mathrm{PL};2}(F,u)\big\|_{\mathcal{L}(H^s,H^{s+r})} &\le K_{r,s}\big(|u|_{C^r_*}\big)|F|_{C^{N_{s+r}+2}}|u|_{C^r_*} \end{aligned}$$

*Moreover, the operator $\mathcal{R}_{\mathrm{PL};2}\big(F,u\big) \in \mathcal{L}(H^s, H^{s+r})$ depends continuously on $u \in H^s \cap C^r_*$ in the operator norm.*

*If in addition, $F = F(x,z)$ is of better regularity $C^{N_{s+r}+3}(\mathbb{T}^n\times\mathbb{R}^L)$ instead of merely $C^{N_{s+r}+2}(\mathbb{T}^n\times\mathbb{R}^L)$, then both operators $T_{F'_z(x,u)} \in \mathcal{L}(H^s,H^s)$ and $\mathcal{R}_{\mathrm{PL};2}\big(F(x,\cdot),u\big) \in \mathcal{L}(H^s,H^{s+r})$ are continuously differentiable in $u \in H^s \cap C^r_*$ with respect to operator norms.*

*Proof.* The proof is a refinement of the standard telescope series argument; see for example, Chapter 10 of [45]. We write

$$\begin{aligned} F(x,u) - F(x,0) &= F(x,S_0u) + \sum_{j=1}^{\infty} F(x,S_ju) - F(x,S_{j-1}u) \\ &= \Big(\int_0^1 F'_z(x,\tau\Delta_0u)\,\mathrm{d}\tau\Big)\Delta_0u + \sum_{j=1}^{\infty}\Big(\int_0^1 F'_z\big(x,S_{j-1}u+\tau\Delta_ju\big)\,\mathrm{d}\tau\Big)\Delta_ju. \end{aligned} \tag{3.6}$$

Defining the multipliers

$$\begin{aligned} m^1_j &= (1-S_{j-3})\big(F'_z(x,0) - \mathrm{Avg}\, F'_z(x,0)\big), \quad j \ge 0 \\ m^2_0 &= \int_0^1 \Big[F'_z\big(x,\tau\Delta_0u\big) - F'_z(x,0)\Big]\,\mathrm{d}\tau - S_{-3}\big(F'_z(x,u) - F'_z(x,0)\big), \\ m^2_j &= \int_0^1 \Big[F'_z\big(x,S_{j-1}u+\tau\Delta_ju\big) - F'_z(x,0)\Big]\,\mathrm{d}\tau - S_{j-3}\big(F'_z(x,u) - F'_z(x,0)\big), \quad j \ge 1 \end{aligned}$$

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

we can rewrite the telescope series (3.6) as

$$F(x,u) - F(x,0) = T_{F_z'(x,u)}u + \sum_{j=0}^{\infty} m_j^1 \Delta_j u + \sum_{j=0}^{\infty} m_j^2 \Delta_j u.$$

Let us then just define, given $u \in H^s \cap C_*^r$, the linear operators

$$\mathcal{R}_{\mathrm{PL};1}(F)\cdot v := \sum_{j=0}^{\infty} m_j^1 \Delta_j v, \quad \mathcal{R}_{\mathrm{PL};2}(F,u)\cdot v := \sum_{j=0}^{\infty} m_j^2 \Delta_j v.$$

In view of Proposition 3.1, it suffices to estimate the two sets of Meyer multipliers $\{m_j^1\}$ and $\{m_j^2\}$.

For the multipliers $\{m_j^1\}$, we estimate them by the simple decay property in Proposition 3.2 for the Zygmund class $C_*^r$. Since $F_z'(x,0) \in C^{N_{s+r}+1}$ by assumption, we obtain by Proposition 3.2 that

$$|\partial^\alpha m_j^1|_{L^\infty} \le C_{s,\alpha} 2^{j(|\alpha|-r)} \big|F_z'(x,0) - \operatorname{Avg} F_z'(x,0)\big|_{C^r}, \quad |\alpha| \le N_{s+r}.$$

Proposition 3.1 then ensures the estimate for the operator $\mathcal{R}_{\mathrm{PL};1}(F)$.

The estimate for multipliers $\{m_j^2\}$ follows by the same idea but is technically more involved. In view of Proposition 3.1, it suffices to show that

$$|\partial^\alpha m_j^2|_{L^\infty} \le K_{s,\alpha}\big(|u|_{C_*^r}\big) 2^{j(|\alpha|-r)} |F|_{C^{N_{s+r}+2}} |u|_{C_*^r}, \quad |\alpha| \le N_{s+r}. \tag{3.7}$$

To obtain (3.7) with $\alpha = 0$, we notice that for $j \ge 1$,

$$\begin{aligned} m_j^2 &= \int_0^1 \Big[\big(F_z'(x, S_{j-1}u + \tau\Delta_j u) - F_z'(x,u)\big)\Big]\,\mathrm{d}\tau + (1 - S_{j-3})\big(F_z'(x,u) - F_z'(x,0)\big) \\ &= \int_{[0,1]^2} F_{zz}''\big(x, \tau_2(S_{j-1}u + \tau_1\Delta_j u)\big)\big((1-S_{j-1})u + \tau_1\Delta_j u\big)\,\mathrm{d}\tau_1\,\mathrm{d}\tau_2 \\ &\quad + (1 - S_{j-3})\int_0^1 F_{zz}''(x,\tau u)u\,\mathrm{d}\tau. \end{aligned} \tag{3.8}$$

The first term in the right-hand-side of (3.8) is directly seen (by Proposition 3.2) to be controlled by $2^{-jr}|F|_{C^2}|u|_{C_*^r}$. For the second term, we simply notice that $F \in C^{N_{s+r}+2} \subset C^{r+2}$, so

$$\big|F_{zz}''(x,\tau u)u\big|_{C_*^r} \le K_s\big(|u|_{C_*^r}\big)|F|_{C^{r+2}}|u|_{C_*^r},$$

which then implies the desired estimate by implementing Proposition 3.2 once again. The estiamte for $m_0^2$ is similar.

To obtain (3.7) with $|\alpha| = N_{s+r}$, we rewrite $m_j^2$ in a different manner: for $j \ge 1$,

$$m_j^2 = \int_{[0,1]^2} F_{zz}''\big(x, \tau_2(S_{j-1}u + \tau_1\Delta_j u)\big)\big(S_{j-1}u + \tau_1\Delta_j u\big)\,\mathrm{d}\tau_1\,\mathrm{d}\tau_2 - S_{j-3}\int_0^1 F_{zz}''(x,\tau u)u\;\mathrm{d}\tau. \tag{3.9}$$

We then use $\big|F_{zz}''(x,\tau u)u\big|_{C_*^r} \le K_s\big(|u|_{C_*^r}\big)|F|_{C^{r+2}}|u|_{C_*^r}$ again: for $|\alpha| = N_{s+r}$, by Proposition 3.2, the second term in (3.9) satisfies

$$\left|\partial^\alpha S_{j-3}\int_0^1 F_{zz}''(x,\tau u)u\;\mathrm{d}\tau\right| \le 2^{j(N_{s+r}-r)} K_s\big(|u|_{C_*^r}\big)|F|_{C^{r+2}}|u|_{C_*^r}.$$

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

As for the first term in (3.9), by the Faà di Bruno formula, the partial derivative

$$\partial^\alpha \Big[ F''_{zz}\big(x, \tau_2(S_{j-1}u + \tau_1 \Delta_j u)\big)\big(S_{j-1}u + \tau_1 \Delta_j u\big)\Big]$$

is a linear combination of terms of the form

$$(\tau_2 D_z)^{2+I} F\big(x, \tau_2(S_{j-1}u + \tau_1 \Delta_j u)\big) \cdot \big[\partial^{\beta_1}\big(S_{j-1}u + \tau_1 \Delta_j u\big)\big]^{i_1} \cdots \big[\partial^{\beta_l}\big(S_{j-1}u + \tau_1 \Delta_j u\big)\big]^{i_l}, \tag{3.10}$$

where the scalar indices $i_1, \cdots, i_l, I \le N_{s+r}$ and multi-indices $\beta_1, \cdots, \beta_l$ satisfy

$$N_{s+r} = i_1|\beta_1| + \cdots + i_l|\beta_l|, \quad I = i_1 + \cdots + i_l.$$

By the Littlewood-Paley characterization of Zygmund functions, such a term has an upper bound

$$C_s|F|_{C^{N_{s+r}+2}} \prod_{p:\beta_p > r} \big(|u|_{C^r_*} 2^j\big)^{i_p(|\beta_p|-r)} \prod_{p:\beta_p \le r} \big(|u|_{C^r_*} j\big)^{i_p} \le K_s\big(|u|_{C^r_*}\big)|F|_{C^{N_{s+r}+2}}|u|_{C^r_*} 2^{j(N_{s+r}-r)}. \tag{3.11}$$

Note that the expression vanishes at least linearly as $u \mapsto 0$ since at least one $i_p \neq 0$. The estimate for $\partial^\alpha m_0^2$ is similar.

Therefore, (3.7) holds for $|\alpha| = N_{s+r}$. By interpolation, it remains valid for those multi-indices $\alpha$ with length between 0 and $N_{s+r}$. This proves the operator norm estimates of the paradifferential remainders.

To prove continuous dependence on $u \in H^s \cap C^r_*$ of the operator $\mathcal{R}_{\mathrm{PL};2}\big(F(x,\cdot), u\big) \in \mathcal{L}(H^s, H^{s+r})$, in view of Proposition 3.1, it suffices to show that when $u_1, u_2$ are close in $C^r_*$, then the Meyer multipliers $m_j^2$ corresponding to $u_1$ and $u_2$ will be close to each other, that is, for $|\alpha| \le N_{s+r}$, the quantities

$$2^{-j(|\alpha|-r)}\big|\partial^\alpha\big(m_j^2(u_1) - m_j^2(u_2)\big)\big|_{L^\infty} \tag{3.12}$$

will be close to zero. This is easily verified if we take into account the assumption $F \in C^{N_{s+r}+2}$. For example, in the summand (3.10), the worst scenario is when all the indices $i_p \equiv 1$, so that (3.10) becomes a "highest order derivative":

$$(\tau_2 D_z)^{2+N_{s+r}} F\big(x, \tau_2(S_{j-1}u + \tau_1 \Delta_j u)\big) \cdot \big[D\big(S_{j-1}u + \tau_1 \Delta_j u\big)\big]^{\otimes N_{s+r}}.$$

When evaluated at $u_1, u_2 \in C^r_*$ that are close in the $C^r_*$ topology, we find that

$$\big[(\tau_2 D_z)^{2+N_{s+r}} \Delta_{12} F\big(x, \tau_2(S_{j-1}u + \tau_1 \Delta_j u)\big)\big]$$

will be close to zero since $D_z^{2+N_{s+r}} F(x,z)$ is continuous, while using Proposition 3.2, we estimate

$$\Big|\Delta_{12}\big[D\big(S_{j-1}u + \tau_1 \Delta_j u\big)\big]^{\otimes N_{s+r}}\Big|_{L^\infty}$$

similarly as in (3.11), with a factor $|\Delta_{12}u|_{C^r_*}$. In conclusion, we obtain that (3.12) is close to zero when $u_1$ is close to $u_2$ in $C^r_*$.

Finally, let us suppose $F$ has better regularity, namely $C^{N_{s+r}+3}(\mathbb{T}^n \times \mathbb{R}^L)$. Continuous differentiability of the operator $\mathcal{R}_{\mathrm{PL};2}\big(F(x,\cdot), u\big)$ in $u$ is a direct consequence of the equality

$$(D_u m_j^2)v = \int_0^1 \Big[F''_{zz}\big(x, S_{j-1}u + \tau\Delta_j u\big)(S_{j-1} + \tau\Delta_j)v\Big]\,\mathrm{d}\tau - S_{j-3}\big(F''_{zz}(x,u)v\big)$$

and Remark 3.1. □

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

## 4. Existence of invariant torus

In this section, we present a detailed proof of Theorem 1.1–1.2 together with some possible extensions.

### *4.1. Algebraic set-up*

Let us recall that we are working with the phase space $\mathbb{T}^n \times \mathbb{R}^n$. We are interested in Hamiltonian functions $h(x, y)$ as in (1.2) and embeddings of $\mathbb{T}^n$ into the phase space close to the "flat" one (1.1)[4]. The torus $\{y = 0\}$ is supposed to be "approximately invariant" under the phase flow of Hamilton vector field $X_h$, and the restriction of $X_h$ is approximately the rotation generated by a constant frequency vector field $\omega \in \mathbb{R}^n$.

Let us define the right inverse of $\nabla_\omega$ as follows:

$$\nabla_\omega^{-1} f(\theta) := \sum_{k \in \mathbb{Z}^n \setminus \{0\}} \frac{\hat{f}(k) e^{ik\cdot\theta}}{i(k \cdot \omega)}.$$

Clearly, when restricted to the subspace of vanishing average, the operator $\nabla_\omega^{-1}$ is also the left inverse of $\nabla_\omega$.

The Diophantine condition (1.3) of $\omega$ enables us to estimate the operator bound of $\nabla_\omega^{-1}$:

$$\left\|\nabla_\omega^{-1} f\right\|_{H^s} \le \gamma^{-1} \|f\|_{H^{s+\tau}}, \quad \text{if } \hat{f}(0) = 0. \tag{4.1}$$

We now study the conjugacy equations (1.6) and (1.9), that is,

$$X_h(u) - \nabla_\omega u = 0$$

and

$$X_{h_\xi}(u) - \nabla_\omega u = 0.$$

To employ the general heuristics discussed in Subsection 2.6, we define a mapping

$$\mathscr{F}(h, u) := X_h(u) - \nabla_\omega u, \quad \text{whose value we denote as } E. \tag{4.2}$$

It is the "error function" measuring whether $u$ is indeed an invariant torus of $X_h$ or not. The problem is then reduced to looking for zero $u \simeq \zeta_0$ of $\mathscr{F}(h, u)$ under the condition of Theorem 1.1, or looking for $\xi \in \mathbb{R}^n$ and zero $u \simeq \zeta_0$ of $\mathscr{F}(h_\xi, u)$ under the condition of Theorem 1.2.

We claim that, to solve $\mathscr{F}(h, u) = 0$, it suffices to solve a slightly "softer" equation

$$\mathscr{F}(h, u) + \begin{pmatrix} 0 \\ \mu \end{pmatrix} = 0 \tag{4.3}$$

for an auxiliary constant vector $\mu \in \mathbb{R}^n$. We directly quote the following geometric lemma from [10]:

**Lemma 4.1** ([10], Lemma 3). *There holds*

$$\mu = \mathrm{Avg}\left( (\partial u^y)^\mathsf{T} \mathscr{F}^x(h, u) - (\partial u^x)^\mathsf{T} \left( \mathscr{F}^y(h, u) - \mu \right) \right).$$

*In particular,* $\mathscr{F}(h, u) = \begin{pmatrix} 0 \\ \mu \end{pmatrix}$ *necessarily implies* $\mu = 0$.

[4] Unfortunately, we have to use $\theta \in \mathbb{T}^n$ to denote the variable on the "abstract torus" since $x$ has already been occupied to label points in phase space.

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

In this subsection, our main task is to compute the linearization of $\mathscr{F}(h,u)$ at a given $u$, as in [59]. For convenience in notation, write

$$A[u] := (DX_h)(u) = \begin{pmatrix} D_x\nabla_y h(u) & D_y\nabla_y h(u) \\ -D_x\nabla_x h(u) & -D_y\nabla_x h(u) \end{pmatrix} \in \boldsymbol{M}_{2n\times 2n}, \tag{4.4}$$

so that

$$D_u\mathscr{F}(h,u)v = A[u]v - \nabla_\omega v.$$

We also write

$$N[u] = \big((\partial u)^{\mathsf{T}}\cdot\partial u\big)^{-1} \in \boldsymbol{M}_{n\times n}, \quad M[u] = \big(\partial u \quad (J\partial u)\cdot N[u]\big) \in \boldsymbol{M}_{2n\times 2n}, \tag{4.5}$$

provided that $\partial u^{\mathsf{T}}\cdot\partial u$ is invertible. This is true when $u$ is $C^1$ close to $\zeta_0$, whence $N[u] = I_n + O\big(\partial(u-\zeta_0)\big)$. This furthermore ensures that

$$M[u] = \begin{pmatrix} I_n & \\ & -I_n \end{pmatrix} + O\big(\partial(u-\zeta_0)\big), \quad M[u]^{-1} = \begin{pmatrix} I_n & \\ & -I_n \end{pmatrix} + O\big(\partial(u-\zeta_0)\big).$$

We then introduce a function $L[u]$, reflecting the "lack of being Lagrangian" for the submanifold $u(\mathbb{T}^n) \subset \mathbb{T}^n\times\mathbb{R}^n$, denoted by

$$L[u] = (\partial u)^{\mathsf{T}}\cdot J\cdot\partial u \in \boldsymbol{M}_{n\times n}. \tag{4.6}$$

Then $L[u]$ is nothing but the matrix representation of the pull-back of the symplectic form on $\mathbb{T}^n\times\mathbb{R}^n$. The submanifold $u(\mathbb{T}^n)\subset\mathbb{T}^n\times\mathbb{R}^n$ is Lagrangian if and only if $L[u]=0$. In particular, this is the case when $u$ does solve the conjugacy equation $\mathscr{F}(h,u)=0$.

The linearization formula for $\mathscr{F}(h,u)$. is summarized as the following lemma[5]:

**Lemma 4.2.** *Suppose $u$ is close to the trivial embedding $\zeta_0$. Let $N[u], M[u], L[u]$ be as in (4.5)(4.6). Define the $n\times n$ matrix function*

$$S[u] = \big(I_n + (N[u]L[u])^2\big)^{-1}N[u]\cdot(\partial u)^{\mathsf{T}}\cdot\big[A[u],J\big]\cdot(\partial u)\cdot N[u]. \tag{4.7}$$

*Then the linearization of $\mathscr{F}(h,u)$ with respect to $u$ satisfies*

$$D_u\mathscr{F}(h,u)\big(M[u]v\big) = M[u]\begin{pmatrix} 0_n & S[u] \\ 0_n & 0_n \end{pmatrix}v - M[u]\nabla_\omega v + \boldsymbol{B}\big(\partial\mathscr{F}(h,u), L[u]\big)v. \tag{4.8}$$

*Here in (4.8), the matrix function $\boldsymbol{B}(\partial E, L)$ is rational[6] in $\partial E$ and $L$, vanishing linearly as $\partial E, L\to 0$, with coefficients being rational functions of $A[u]$, $\partial u$ and $\partial^2 u$. We may then rewrite (4.8) as the equivalent form*

$$D_u\mathscr{F}(h,u)v = M[u]\begin{pmatrix} 0_n & S[u] \\ 0_n & 0_n \end{pmatrix}M[u]^{-1}v - M[u]\nabla_\omega\big(M[u]^{-1}v\big) + \boldsymbol{B}\big(\partial\mathscr{F}(h,u), L[u]\big)M[u]^{-1}v, \tag{4.9}$$

*Proof.* Throughout this proof, for simplicity in notation, we shall drop the $[u]$ (indicating dependence on $u$), write, for example, $A$ for $A[u]$, and denote $E := \mathscr{F}(h,u)$. For a given $u$ and its increment $v$, we

[5]The authors would like to thank Giovanni Forni and Nicolò Tedesco for pointing out a computational error in the original version of the paper and for subsequent discussions on the details of the correction.

[6]We say that a matrix function $f(X)$ is rational in the argument $X$, if the entries of $f(X)$ are all rational functions of entries of $X$.

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

first write

$$\begin{aligned} D_u\mathscr{F}(h,u)(M[u]v) &= AMv - (\nabla_\omega M)v - M\nabla_\omega v \\ &=: \begin{pmatrix} C_1 & C_2 \end{pmatrix} v - M\nabla_\omega v \end{aligned}$$

Here we write the matrix $AM - (\nabla_\omega M) \in \boldsymbol{M}_{2n\times 2n}$ into block form $\begin{pmatrix} C_1 & C_2 \end{pmatrix}$. The task is to determine the blocks $C_1, C_2 \in \boldsymbol{M}_{2n\times n}$.

It is fairly easy to see that $C_1 = \partial E$. In fact, we may write $M[u]$ into block form as

$$M = \begin{pmatrix} \partial u & (J\partial u)N \end{pmatrix} =: \begin{pmatrix} M_1 & M_2 \end{pmatrix}.$$

Then obviously

$$\begin{aligned} C_1 &= AM_1 - \nabla_\omega M_1 \\ &= A\partial u - \nabla_\omega \partial u = \partial\big(\mathscr{F}(h,u)\big) = \partial E. \end{aligned} \tag{4.10}$$

The true difficulty is with the expression of $C_2 = A(J\partial u)N - \nabla_\omega\big((J\partial u)N\big)$. We try to find $n\times n$ matrices $Q, W$, such that

$$C_2 = (\partial u)Q + (J\partial u)NW = M_1 Q + M_2 W. \tag{4.11}$$

In solving this matrix system, we take into account that $L = (\partial u)^{\mathsf{T}} J\partial u$ is supposed to be close to 0. Multiplying (4.11) from left by $(\partial u)^{\mathsf{T}} J$, using in sequence the Hamiltonian nature $JA = -A^{\mathsf{T}} J$ and the symplectic nature $J^2 = -I_{2n}$, we obtain

$$\begin{aligned} LQ - W &= (\partial u)^{\mathsf{T}} JA(J\partial u)N - (\partial u)^{\mathsf{T}} J\nabla_\omega\big((J\partial u)N\big) \\ &= (\partial u)^{\mathsf{T}} A^{\mathsf{T}}(\partial u)N + (\partial u)^{\mathsf{T}}(\nabla_\omega \partial u)N + (\partial u)^{\mathsf{T}}(\partial u)\nabla_\omega N \\ &= (\partial E)^{\mathsf{T}}(\partial u)N. \end{aligned} \tag{4.12}$$

Here in the last step we differentiated the definition $(\partial u)^{\mathsf{T}}(\partial u)N = I_n$ to deduce

$$(\partial u)^{\mathsf{T}}(\nabla_\omega \partial u)N + (\partial u)^{\mathsf{T}}(\partial u)\nabla_\omega N = -\big[\nabla_\omega (\partial u)^{\mathsf{T}}\big](\partial u)N,$$

and then use the last equality of (4.10).

On the other hand, multiplying (4.11) from left by $N(\partial u)^{\mathsf{T}}$, we obtain

$$\begin{aligned} Q + NLNW &= N(\partial u)^{\mathsf{T}} A\cdot (J\partial u)N - N(\partial u)^{\mathsf{T}}\nabla_\omega\big((J\partial u)N\big) \\ &= N(\partial u)^{\mathsf{T}}\Big[A, J\Big](\partial u)N + N(\partial u)^{\mathsf{T}} JA(\partial u)N \\ &\quad - N(\partial u)^{\mathsf{T}} J(\nabla_\omega \partial u)N - N(\partial u)^{\mathsf{T}} J(\partial u)\nabla_\omega N \\ &= N(\partial u)^{\mathsf{T}}\Big[A, J\Big](\partial u)N + N(\partial u)^{\mathsf{T}} J(\partial E)N - NL\nabla_\omega N. \end{aligned} \tag{4.13}$$

We can then directly solve from (4.12)(4.13) the expression of $Q, W$:

$$\begin{aligned} Q &= S + \big(I_n + (NL)^2\big)^{-1}\Big(NLN(\partial E)^{\mathsf{T}}(\partial u)N + N(\partial u)^{\mathsf{T}} J(\partial E)N - NL\nabla_\omega N\Big) \\ W &= LQ - (\partial E)^{\mathsf{T}}(\partial u)N. \end{aligned} \tag{4.14}$$

Therefore, we immediately see that both $W - LS$ and $Q - S$ are (matrix) rational functions of $L$ and $\partial E$, vanishing linearly in $L$ and $\partial E$, with coefficients being rational functions of $\partial u$ and $\partial^2 u$.

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

Combining (4.10) and (4.11), we find that the matrix $AM - (\nabla_\omega M)$ has block-form expression

$$\begin{aligned} AM - (\nabla_\omega M) &= \begin{pmatrix} \partial E & M_1 S \end{pmatrix} + \begin{pmatrix} 0_{2n\times n} & M_1(Q-S) + M_2 W \end{pmatrix} \\ &= M \begin{pmatrix} 0_n & S \\ 0_n & 0_n \end{pmatrix} + \begin{pmatrix} \partial E & M_1(Q-S) + M_2 W \end{pmatrix}. \end{aligned}$$

Therefore, the remainder $\boldsymbol{B}(\partial E, L) = \begin{pmatrix} \partial E & M_1(Q-S) + M_2 W \end{pmatrix}$, which, by (4.14), is a matrix rational function of $L$ and $\partial E$, vanishing linearly in $L$ and $\partial E$, with coefficients being rational functions of $A$, $\partial u$ and $\partial^2 u$. □

**Remark 4.1.** Lemma 4.2 is in fact equivalent to Lemma 20 of [59]. It also appeared in [10]. However, we are not able to find an explicit expression for $D_u \mathscr{F}(h,u)v$ in these references, which is the reason that we choose to write down our own proof. We observe that (4.9) plays the same role as (2.7).

A key feature of the Lagrangian character $L[u]$ is that it is in proportion to the error $E = \mathscr{F}(h,u)$. In fact by Lemma 19 of [59] or Lemma 5 of [10], if we define the matrix function

$$L_1(u,E) := \partial\big((\partial u)^{\mathsf{T}} \cdot JE\big) - \Big[\partial\big((\partial u)^{\mathsf{T}} \cdot JE\big)^{\mathsf{T}}\Big)\Big]^{\mathsf{T}},$$

then there holds

$$\nabla_\omega L[u] = L_1(u, \mathscr{F}(h,u)) \quad \text{or equivalently} \quad L[u] = \nabla_\omega^{-1} L_1(u, \mathscr{F}(h,u)). \tag{4.15}$$

Note that due to the presence of $\partial$, the function $L_1(u,E)$ necessarily has average zero. In other words, the matrix function $\nabla_\omega L[u]$ is nothing but the matrix representation of the exterior differential of the 1-form represented by $(\partial u)^{\mathsf{T}} \cdot J\mathscr{F}(h,u)$. Of course, this is a refined version of the observation that an invariant torus is necessarily a Lagrangian submanifold. Therefore, the mapping $\boldsymbol{B}\big(\mathscr{F}(h,u), L[u]\big)$ in (4.9) vanishes linearly in $E$ if $u$ is approximately invariant:

**Corollary 4.2.1.** *Suppose $\omega$ satisfies the Diophantine condition (1.3), so that (4.1) holds. Suppose $u \in H^s$ for some $s > \max(\tau+2, n/2+2)$, while $|u-\zeta_0|_{C^2}$ is sufficiently close to zero. Then for any constant shift $(0,\mu)^{\mathsf{T}} \in \mathbb{R}^n \times \mathbb{R}^n$, there holds*

$$\big\|\boldsymbol{B}\big(\partial\mathscr{F}(h,u), L[u]\big)\big\|_{H^{s-(\tau+2)}} \le \gamma^{-1} K_s\big(|h|_{C^{[s]+3}}, \|u\|_{H^s}\big)\big\|\mathscr{F}(h,u) + (0,\mu)^{\mathsf{T}}\big\|_{H^{s-1}}.$$

*The function $K_s$ is increasing in both arguments and does not depend on the shift $(0,\mu)^{\mathsf{T}} \in \mathbb{R}^n \times \mathbb{R}^n$.*

*Proof.* By Lemma 4.2, we know that $\boldsymbol{B}\big(\partial\mathscr{F}(h,u), L[u]\big)$ is a rational function of $\partial\mathscr{F}(h,u)$ and $L[u]$, vanishing linearly as the arguments $\to 0$, with coefficients being rational functions of $A[u], \partial u, \partial^2 u$. Since $|u-\zeta_0|_{C^2}$ is close to 0, we find that both $\partial\mathscr{F}(h,u)$ and $L[u]$ are $C^0$ close to 0, since they only involve $u$ and $\partial u$. Consequently, for example, the matrix $I_n + (N[u]L[u])^2$ is $H^{s-1}$ close to the identity matrix.

Obviously, shifting $\mathscr{F}(h,u)$ by a constant addendum does not affect the value of $\boldsymbol{B}\big(\partial\mathscr{F}(h,u), L[u]\big)$. Therefore,

$$\big\|\boldsymbol{B}\big(\partial\mathscr{F}(h,u), L[u]\big)\big\|_{H^{s-(\tau+2)}} \le K_s\big(|h|_{C^{[s]+3}}, \|u\|_{H^s}\big)\Big(\big\|\mathscr{F}(h,u) + (0,\mu)^{\mathsf{T}}\big\|_{H^{s-1}} + \|L[u]\|_{H^{s-(\tau+2)}}\Big).$$

Now we make use of (4.15). Note that for the matrix function

$$L_1(u,E) = \partial\big((\partial u)^{\mathsf{T}} \cdot JE\big) - \Big[\partial\big((\partial u)^{\mathsf{T}} \cdot JE\big)^{\mathsf{T}}\Big)\Big]^{\mathsf{T}},$$

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

shifting $E$ by a constant addendum $(0,\mu)^{\mathsf{T}}$ yields the same value of $L_1(u,E)$, by a direct computation (using $\partial_i\partial_j u = \partial_j\partial_i u$). Therefore, by (4.15), we have

$$\begin{aligned}\|L[u]\|_{H^{s-(\tau+2)}} &\le \gamma^{-1}\big\|L_1\big(u,\mathscr{F}(h,u)+(0,\mu)^{\mathsf{T}}\big)\big\|_{H^{s-2}}\\ &\lesssim \gamma^{-1}K_s\big(\|u\|_{H^s}\big)\big\|\mathscr{F}(h,u)+(0,\mu)^{\mathsf{T}}\big\|_{H^{s-1}}.\end{aligned}$$

This concludes the proof. □

### *4.2. Deriving parahomological equation*

So far we have been mostly quoting and rewriting the algebraic findings reported in Section 8 of [59]. However, from this point, we will proceed quite differently from [59], or any other known reference on KAM theory, for example, [10] – instead of a Newtonian algorithm involving approximate right inverse, we employ the idea of Subsection 2.6 to directly solve the parahomological equation. We may treat Theorem 1.1 and 1.2 in a unified manner, since for Theorem 1.1 it suffices to require that the parameter $\xi$ must be zero.

So we restrict to a neighbourhood of the already known approximate solution $\zeta_0$. We notice that $\mathscr{F}(h_\xi,u)$ is a first order nonlinear partial differential operator acting on the unknown $(u,\xi)\in H^s\times\mathbb{R}^n$, and in fact no differential in the constant $\xi$ is involved at all. Let us recall from (1.4) and (1.7) that $e_0=\mathscr{F}(h,\zeta_0)$. Then the paralinearization formula

$$\mathscr{F}(h_\xi,u)+\begin{pmatrix}0\\ \mu\end{pmatrix}=e_0+\begin{pmatrix}\xi\\ \mu\end{pmatrix}+T_{D_u\mathscr{F}(h,u)}(u-\zeta_0)+\mathcal{R}_{\mathrm{PL}}(X_h,u-\zeta_0)\cdot(u-\zeta_0) \tag{4.16}$$

is valid. The expression $\boldsymbol{B}(E,L[u])$ involves only $\partial E$, so it does not change if $E$ is shifted by a constant vector. Thus, writing for simplicity

$$E:=\mathscr{F}(h_\xi,u)+\begin{pmatrix}0\\ \mu\end{pmatrix},$$

we actually have[7], by Lemma 4.2, the following equality:

$$\begin{aligned}T_{D_u\mathscr{F}(h,u)}(u-\zeta_0)&=\mathbf{Op}^{\mathrm{PM}}\left[M[u]\begin{pmatrix}0_n & T_{S[u]}\\ 0_n & 0_n\end{pmatrix}M[u]^{-1}\right](u-\zeta_0)\\ &\quad-\mathbf{Op}^{\mathrm{PM}}\Big(M[u]\big(\nabla_\omega M[u]^{-1}\big)+\nabla_\omega\Big)(u-\zeta_0)+\mathbf{Op}^{\mathrm{PM}}\Big(\boldsymbol{B}(E,L[u])M[u]^{-1}\Big)(u-\zeta_0).\end{aligned}$$

Thus the paralinearization formula (4.16) becomes

$$\begin{aligned}E=e_0+\begin{pmatrix}\xi\\ \mu\end{pmatrix}&+T_{M[u]}\begin{pmatrix}0_n & T_{S[u]}\\ 0_n & 0_n\end{pmatrix}T_{M[u]^{-1}}(u-\zeta_0)-T_{M[u]}\nabla_\omega T_{M[u]^{-1}}(u-\zeta_0)\\ &+\mathbf{Op}^{\mathrm{PM}}\Big(\boldsymbol{B}(E,L[u])M[u]^{-1}\Big)(u-\zeta_0)+\mathcal{R}_1[u](u-\zeta_0)+\mathcal{R}_{\mathrm{PL}}(X_h,u-\zeta_0)\cdot(u-\zeta_0).\end{aligned} \tag{4.17}$$

Here the linear operator

$$\begin{aligned}\mathcal{R}_1[u]=\mathbf{Op}^{\mathrm{PM}}&\left[M[u]\begin{pmatrix}0_n & T_{S[u]}\\ 0_n & 0_n\end{pmatrix}M[u]^{-1}\right]-T_{M[u]}\begin{pmatrix}0_n & T_{S[u]}\\ 0_n & 0_n\end{pmatrix}T_{M[u]^{-1}}\\ &-\mathbf{Op}^{\mathrm{PM}}\Big[M[u]\big(\nabla_\omega M[u]^{-1}\big)+\nabla_\omega\Big]+T_{M[u]}\nabla_\omega T_{M[u]^{-1}}\end{aligned} \tag{4.18}$$

is produced in view of Proposition 3.4. Note that we used the Leibniz rule $\partial T_a u = T_{\partial a}u + T_a\partial u$.

[7]Our notation (2.10) is that $\mathbf{Op}^{\mathrm{PM}}(a)=T_a$, and we prefer to use $\mathbf{Op}^{\mathrm{PM}}$ here in case the expression of $a$ is very lengthy.

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

**Lemma 4.3.** *Fix an index $s > 2\tau + 2 + n/2 + \varepsilon$ and set $r = s - n/2$. Suppose the Hamilton function $h \in C^{N_{s+r}+3} u \in H^s(\mathbb{T}^n; \mathbb{T}^n \times \mathbb{R}^n)$ is close to $\zeta_0$ in that space. Then the operator $\mathcal{R}_1[u]$ defined by (4.18) maps $H^t$ to $H^{t+2\tau+\varepsilon}$ for every index $t \in \mathbb{R}$: with $K_{t,s}$ being some increasing function,*

$$\|\mathcal{R}_1[u]v\|_{H^{t+r-2}} \leq K_{t,s}\big(|h|_{C^{N_{s+r}+3}}, \|u\|_{H^s}\big)\|u - \zeta_0\|_{H^s}\|v\|_{H^t}.$$

*Furthermore, as such an operator, $\mathcal{R}_1[u]$ has $C^1$ dependence on $u \in H^s$.*

*Proof.* For simplicity of notation we write

$$a = M[u], \quad b = \begin{pmatrix} 0_n & T_{S[u]} \\ 0_n & 0_n \end{pmatrix}, \quad c = M[u]^{-1}.$$

Then (4.18) is rewritten as

$$\begin{aligned} \mathcal{R}_1[u] &= T_{abc} - T_a T_b T_c + (T_a T_c - 1)\nabla_\omega + T_a T_{\nabla_\omega c} - T_{a\nabla_\omega c} \\ &= \mathcal{R}_{\mathrm{CM}}(ab, c) + \mathcal{R}_{\mathrm{CM}}(a, b)T_c + \mathcal{R}_{\mathrm{CM}}(a, c)\nabla_\omega - \mathcal{R}_{\mathrm{CM}}(a, \nabla_\omega c). \end{aligned}$$

Here we used the fact that $ac$ is the identity matrix. We then recall from (4.5) that $a = \mathrm{diag}(I_n, -I_n) + O\big(\partial(u - \zeta_0)\big)$, and $S[u]$ is given by (4.7). Therefore, we have

$$\big|S[u]\big|_{C^{r-1}_*} \leq K_s\big(|h|_{C^{N_{s+r}}}, \|u\|_{H^s}\big),$$

and

$$\big|a - \mathrm{diag}(I_n, -I_n)\big|_{C^{r-1}_*} + \big|c - \mathrm{diag}(I_n, -I_n)\big|_{C^{r-1}_*} \leq \|u - \zeta_0\|_{H^s}.$$

We then use Proposition 3.4; notice that $\mathcal{R}_{\mathrm{CM}}(\cdot, \cdot)$ does not change its value when the inputs are shifted by constants. Therefore, $\mathcal{R}_1[u]$ is a smoothing operator of order $r - 2$, vanishing linearly as $u \to \zeta_0$. This gives the desired operatorial bound. □

We can now formulate the parahomological equation in the unknown $(u, \xi, \mu)$, where $\xi, \mu \in \mathbb{R}^n$ are constant parameters to be fixed. The equation reads

$$\begin{aligned} &T_{M[u]}\begin{pmatrix} 0_n & T_{S[u]} \\ 0_n & 0_n \end{pmatrix} T_{M[u]^{-1}}(u - \zeta_0) - T_{M[u]}\nabla_\omega T_{M[u]^{-1}}(u - \zeta_0) + \begin{pmatrix} \xi \\ \mu \end{pmatrix} \\ &\quad = -e_0 - \mathcal{R}_1[u](u - \zeta_0) - \mathcal{R}_{\mathrm{PL}}(X_h, u - \zeta_0) \cdot (u - \zeta_0). \end{aligned} \tag{4.19}$$

In other words, it aims to cancel out all terms in the right-hand-side of (4.17) but the one linear in $E$.

To solve the parahomological equation (4.19), we need a lemma concerning *linear* parahomological equations, which is essentially the simple equation $\nabla_\omega v = f$.

**Lemma 4.4.** *Let $\omega$ be a Diophantine frequency vector satisfying (1.3). Set*

$$M_Q = \begin{cases} \max\Big(\big|(\mathrm{Avg}\, Q)^{-1}\big|, |Q|_{L^\infty}\Big) & \text{under the assumption of Theorem 1.1,} \\ |Q|_{L^\infty} & \text{under the assumption of Theorem 1.2.} \end{cases}$$

*There are constants $\rho_1, \rho_2$ depending on $|h|_{C^3}$ and $M_Q$ with the following property. If $|e_0|_{C^1} \leq \rho_1$, and the embedding $u : \mathbb{T}^n \mapsto \mathbb{T}^n \times \mathbb{R}^n$ is such that $|u - \zeta_0|_{C^1} \leq \rho_2$, then the linear parahomological equation in the unknown $(v, \xi, \mu) \in H^s \times \mathbb{R}^n \times \mathbb{R}^n$*

$$T_{M[u]}\begin{pmatrix} 0_n & T_{S[u]} \\ 0_n & 0_n \end{pmatrix} T_{M[u]^{-1}} v - T_{M[u]}\nabla_\omega T_{M[u]^{-1}} v + \begin{pmatrix} \xi \\ \mu \end{pmatrix} = f \tag{4.20}$$

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

*has a linear solution operator*

$$\big(v^x,\, v^y,\, \xi,\, \mu\big) = \big(\mathfrak{L}^x[u]f,\, \mathfrak{L}^y[u]f,\, P^x[u]f,\, P^y[u]f\big),$$

*satisfying the following estimates: with $K_s$ being increasing functions in the arguments,*

$$\begin{aligned}
\big\|\mathfrak{L}^x[u]f\big\|_{H^s} &\le K_s\big(M_Q, |h|_{C^3}\big)\gamma^{-2}\|f\|_{H^{s+2\tau}}\\
\big\|\mathfrak{L}^y[u]f\big\|_{H^s} &\le K_s\big(M_Q, |h|_{C^3}\big)\gamma^{-1}\|f\|_{H^{s+\tau}}\\
|P^x[u]f| + |P^y[u]f| &\le K_s\big(M_Q, |h|_{C^3}\big)\gamma^{-1}\|f\|_{H^{s+\tau}}.
\end{aligned}$$

*Moreover, the four linear operators of concern are all continuously differentiable mappings from $u \in C^1$ to the space of linear operators (with operator norm). In particular, the solution operator $\xi = P^x[u]f$ is fixed as 0 under the assumption of Theorem 1.1.*

*Proof.* Since $M[u] = \operatorname{diag}(I_n, -I_n) + O\big(\partial(u-\zeta_0)\big)$, it follows that $T_{M[u]}$ and $T_{M[u]^{-1}}$ are both invertible bounded operator from $H^s$ to $H^s$ if $|u-\zeta_0|_{C^1}$ is small. Note further that for a constant $\lambda$ and a function $A$, there holds $T_A\lambda = (\operatorname{Avg} A)\lambda$. Thus, writing

$$T^{-1}_{M[u]}\begin{pmatrix}\xi\\ \mu\end{pmatrix} = \begin{pmatrix}\xi_1\\ \mu_1\end{pmatrix}, \quad T^{-1}_{M[u]}f = f_1, \quad T_{M[u]^{-1}}v = v_1 \tag{4.21}$$

for some constant vectors $\xi_1, \mu_1 \in \mathbb{R}^n$ to be determined, the linear parahomological equation (4.20) is equivalent to

$$\begin{pmatrix}0_n & T_{S[u]}\\ 0_n & 0_n\end{pmatrix}v_1 - \nabla_\omega v_1 + \begin{pmatrix}\xi_1\\ \mu_1\end{pmatrix} = f_1.$$

Furthermore, recalling the expression (1.2) of $h(x,y)$ and (4.7), we have

$$S[u] = Q + \partial^2 a_0 + O\big(\partial(u-\zeta_0)\big), \tag{4.22}$$

where the implicit constant depends on $|h|_{C^3}$. Noting that

$$e_0 = \mathscr{F}(h,\zeta_0) = \begin{pmatrix}a_1-\omega\\ \partial a_0\end{pmatrix}, \tag{4.23}$$

it follows that $S[u]$ is $C^0$-close to $Q$ if $|e_0|_{C^1}$ and $|u-\zeta_0|_{C^1}$ are small.

In components of $v_1$, the linear parahomological equation becomes

$$\begin{aligned}
T_{S[u]}v_1^y - \nabla_\omega v_1^x + \xi_1 &= f_1^x,\\
-\nabla_\omega v_1^y + \mu_1 &= f_1^y.
\end{aligned}$$

We discuss how to solve this system under the assumption of Theorem 1.1 and 1.2 separately. For simplicity of notation, we define

$$\mathcal{A}f := f - \operatorname{Avg} f.$$

□

**Case 1:** Avg$Q$ **is invertible**

This is the nondegeneracy condition in Therorem 1.1. We fix $\xi = 0$ in this case, and obtain from (4.21) the following:

$$\begin{pmatrix}0\\ \mu\end{pmatrix} = \begin{pmatrix}\xi_1\\ -\mu_1\end{pmatrix} + O\big(\partial(u-\zeta_0)\big)\begin{pmatrix}\xi_1\\ \mu_1\end{pmatrix}. \tag{4.24}$$

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

We first solve

$$v_1^y = \operatorname{Avg} v_1^y - \nabla_\omega^{-1}\mathcal{A} f_1^y, \quad \mu_1 = \operatorname{Avg} f_1^y, \tag{4.25}$$

where $\operatorname{Avg} v_1^y$ is yet to be determined. With (4.24), we can solve $\xi_1$, thus $\mu$, as linear mappings of $\mu_1$. Since the norm of $T^{-1}_{M[u]}$ from $H^s$ to $H^s$ depends only on $|u-\zeta_0|_{C^1}$, this immediately yields

$$\left\|\mathcal{A} v_1^y\right\|_{H^s} + |\mu| + |\xi_1| \le C_s \gamma^{-1} \|f\|_{H^{s+\tau}}.$$

Here we used (4.1).

We then choose $\operatorname{Avg} v_1^y$ to balance the mean value of the equation along $x$-direction. By assumption $S[\zeta_0]$ is invertible, so by (4.22), the matrix-valued function $S[u] = Q + O(\partial e_0) + O\big(\partial(u-\zeta_0)\big)$ is still invertible if $|e_0|_{C^1}$ and $|u-\zeta_0|_{C^1}$ are small, and the implicit constant here depends on $|h|_{C^3}$. In that case, $\left|(\operatorname{Avg} S[u])^{-1}\right| \lesssim M_Q$. We can then set

$$\operatorname{Avg} v_1^y = \big(\operatorname{Avg} S[u]\big)^{-1} \operatorname{Avg}\big(f_1^x - \xi_1 - T_{S[u]}\mathcal{A} v_1^y\big),$$

and solve

$$v_1^x = \nabla_\omega^{-1}\mathcal{A}\big(T_{S[u]}\mathcal{A} v_1^y + \xi_1 - f_1^x\big).$$

By (4.7), $S[u]$ depends on up to second order derivative of $h$, so the solutions are immediately estimated as

$$\left\|v_1^x\right\|_{H^s} \le K_s\big(M_Q, |h|_{C^3}\big)\gamma^{-2}\|f\|_{H^{s+2\tau}}.$$

Here we used (4.1) once again.

**Case 2: Nontrivial parameter $\xi$**

This case matches with Theorem 1.2, whence no requirement about nondegeneracy of $Q$ is posed. In this case the parameter $\xi$ cannot be fixed as 0. But we still obtain from (4.21) the following:

$$\begin{pmatrix}\xi\\ \mu\end{pmatrix} = \begin{pmatrix}\xi_1\\ -\mu_1\end{pmatrix} + O\big(\partial(u-\zeta_0)\big)\begin{pmatrix}\xi_1\\ \mu_1\end{pmatrix}. \tag{4.26}$$

We then first solve

$$v_1^y = -\nabla_\omega^{-1}\mathcal{A} f_1^y, \quad \mu_1 = \operatorname{Avg} f_1^y, \tag{4.27}$$

that is, the mean value of $v_1^y$ is fixed as 0. For the equation along $x$-direction, we solve

$$v_1^x = -\nabla_\omega^{-1}\mathcal{A}\big(f_1^x - T_{S[u]} v_1^y\big), \quad \xi_1 = \operatorname{Avg}\big(f_1^x - T_{S[u]} v_1^y\big).$$

Using (4.1), they immediately yield the similar estimates as in Case 1.

Finally, continuous differentiability of the operators for $u \in C^1$ follows immediately from Proposition 3.3 and 3.4, that is, continuous differentiability of $T_a$ in $a$ and $T_{ab} - T_a T_b$ in $(a,b)$.

### *4.3. Proof of Theorem 1.1–1.2*

Recall that we set $r = s - n/2$, and $N_{s+r}$ to be the least integer $> s + r$. Since we are working with a perturbative problem, we assume that, for example, $\|u-\zeta_0\|_{H^s} \le 2$, and the requirement of Lemma 4.4

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

is also fulfilled: with $\rho_1 = \rho_1\big(M_Q, |h|_{C^3}\big)$, $\rho_2 = \rho_2\big(M_Q, |h|_{C^3}\big)$ as in Lemma 4.4, we have

$$\|u - \zeta_0\|_{H^s} \le 2, \quad |e_0|_{C^1} \le \rho_1, \quad |u - \zeta_0|_{C^1} \le \rho_2. \tag{4.28}$$

The parahomological equation (4.19) is then converted to the following parainverse equation:

$$\begin{aligned} u - \zeta_0 &= -\begin{pmatrix} \mathfrak{Q}^x[u] - P^x[u] \\ \mathfrak{Q}^y[u] - P^y[u] \end{pmatrix} e_0 \\ &\quad - \begin{pmatrix} \mathfrak{Q}^x[u] - P^x[u] \\ \mathfrak{Q}^y[u] - P^y[u] \end{pmatrix} \big(\mathcal{R}_1[u](u - \zeta_0) - \mathcal{R}_{\mathrm{PL}}(X_h, u - \zeta_0) \cdot (u - \zeta_0)\big) \\ &=: \mathcal{G}_1(u) + \mathcal{G}_2(u) \end{aligned} \tag{4.29}$$

We analyze the right-hand side of (4.29). The requirements (4.28) meet the conditions of Lemma 4.4, so

$$\|\mathcal{G}_1(u)\|_{H^{s+\varepsilon}} \le K_s\big(M_Q, |h|_{C^3}\big)\gamma^2 \|e_0\|_{H^{s+2\tau+\varepsilon}}. \tag{4.30}$$

We turn to $\mathcal{G}_2$. The paraproduct remainder $\mathcal{R}_1[u](u - \zeta_0)$ is estimated by Lemma 4.3:

$$\big\|\mathcal{R}_1[u](u - \zeta_0)\big\|_{H^{s+r-2}} \le K_s\big(|h|_{C^{N_{s+r}}}\big)\|u - \zeta_0\|_{H^s}^2, \tag{4.31}$$

and $\mathcal{R}_1[u](u - \zeta_0)$ depends continuously on $u$ (with respect to Sobolev norms $H^s$ and $H^{s+r-2}$). As for the paralinearization remainder $\mathcal{R}_{\mathrm{PL}}$, by Proposition 3.5, recalling (4.4),

$$\mathcal{R}_{\mathrm{PL}}(X_h, u - \zeta_0) \cdot (u - \zeta_0) = \mathcal{R}_{\mathrm{PL};1}\big(A[\zeta_0]\big)(u - \zeta_0) + \mathcal{R}_{\mathrm{PL};2}(X_h, u - \zeta_0) \cdot (u - \zeta_0). \tag{4.32}$$

Recalling further the definition (1.2) of $h(x, y)$, we find

$$A[\zeta_0] = \begin{pmatrix} \partial a_1(\theta) & Q(\theta) \\ \partial^2 a_0(\theta) & -(\partial a_1)^{\mathsf{T}}(\theta) \end{pmatrix}.$$

On the other hand, recalling the definition (4.23), we find

$$\|a_1 - \omega\|_{H^{s+\tau+\varepsilon}} + \|\partial a_0\|_{H^{s+\tau+\varepsilon}} = \|e_0\|_{H^{s+2\tau+\varepsilon}},$$

that is, $a_0$ is almost a constant, and $a_1$ almost equals the constant vector $\omega$. As a result, we find $A[\zeta_0]$ is almost a constant:

$$\begin{aligned} \big|A[\zeta_0] - \operatorname{Avg} A[\zeta_0]\big|_{C^r_*} &\le C_s\big(\|a_1 - \omega\|_{H^{s+\tau+\varepsilon}} + \|\partial a_0\|_{H^{s+\tau+\varepsilon}} + |Q - \operatorname{Avg} Q|_{C^r_*}\big) \\ &\le C_s\big(\|e_0\|_{H^{s+2\tau+\varepsilon}} + |e_1|_{C^r_*}\big). \end{aligned}$$

Implementing the paralinearization theorem, that is, Proposition 3.5, we estimate (4.32) as (note that $X_h = J\nabla_{x,y} h \in C^{N_{s+r}+2}$)

$$\begin{aligned} &\big\|\mathcal{R}_{\mathrm{PL}}(X_h, u - \zeta_0) \cdot (u - \zeta_0)\big\|_{H^{s+r}} \\ &\quad \le C_s\Big(\big(\|e_0\|_{H^{s+2\tau+\varepsilon}} + |e_1|_{C^r_*}\big)\|u - \zeta_0\|_{H^s} + |h|_{C^{N_{s+r}+3}}\|u - \zeta_0\|_{H^s}^2\Big). \end{aligned} \tag{4.33}$$

Furthermore, by Proposition 3.5, the operator $\mathcal{R}_{\mathrm{PL}}\big(X_h(\zeta_0 + \cdot), u - \zeta_0\big) \in \mathcal{L}(H^s, H^{s+r})$ depends continuously on $u \in H^s$.

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

Recall that by assumption, $s+r-2 = 2s-n/2-2 \geq s+2\tau+\varepsilon$. Combining (4.30)(4.31)(4.33), using Lemma 4.4, we conclude that if (4.28) holds, then

$$\begin{aligned}&\|\mathcal{G}_1(u)\|_{H^{s+\varepsilon}} + \|\mathcal{G}_2(u)\|_{H^{s+\varepsilon}}\\ &\quad\leq K_s\big(M_Q, |h|_{C^{N_{s+r+3}}}\big)\gamma^{-2}\Big(\|e_0\|_{H^{s+2\tau+\varepsilon}} + \big(\|e_0\|_{H^{s+2\tau+\varepsilon}} + |e_1|_{C^r_*}\big)\|u-\zeta_0\|_{H^s} + \|u-\zeta_0\|^2_{H^s}\Big).\end{aligned} \tag{4.34}$$

By some elementary algebra with quadratic polynomials, there are decreasing functions $c_0, c_1$ and an increasing function $c_2$, with the following property: if

$$\|e_0\|_{H^{s+2\tau+\varepsilon}} \leq \gamma^4 c_0\big(M_Q, |h|_{C^{N_{s+r+3}}}\big), \quad |e_1|_{C^r_*} \leq \gamma^2 c_1\big(M_Q, |h|_{C^{N_{s+r+3}}}\big),$$

then the positive number $\rho = \|e_0\|_{H^{s+\tau+\varepsilon}}\gamma^2 c_2\big(M_Q, |h|_{C^{N_{s+r+3}}}\big)$ satisfies the quadratic inequality

$$K_s\big(M_Q, |h|_{C^{N_{s+r+3}}}\big)\gamma^{-2}\Big(\|e_0\|_{H^{s+2\tau+\varepsilon}} + \big(\|e_0\|_{H^{s+2\tau+\varepsilon}} + |e_1|_{C^r_*}\big)\rho + \rho^2\Big) \leq \rho.$$

Here $K_s$ is the increasing function in (4.34). In other words,

$$\|\mathcal{G}_1(u)\|_{H^{s+\varepsilon}} + \|\mathcal{G}_2(u)\|_{H^{s+\varepsilon}} \leq \rho \quad \text{if } \|u-\zeta_0\|_{H^s} \leq \rho \text{ and (4.28) holds.}$$

In that case, the mapping $\zeta_0+\mathcal{G}_1+\mathcal{G}_2$ in the right-hand-side of (4.29) maps the closed ball $\bar{B}_\rho(\zeta_0) \subset H^s$ to itself. Furthermore it is continuous with respect to $u \in \bar{B}_\rho(\zeta_0) \subset H^s$ and the image is totally bounded (since the embedding $H^{s+\varepsilon} \subset H^s$ is compact). By the Schauder fixed point theorem, the mapping $u \mapsto \zeta_0 + \mathcal{G}_1(u) + \mathcal{G}_2(u)$ has a fixed point $u \in \bar{B}_\rho(\zeta_0)$, and in fact

$$\|u-\zeta_0\|_{H^s} \leq c_2\big(M_Q, |h|_{C^{N_{s+r+3}}}\big)\gamma^{-2}\|e_0\|_{H^{s+2\tau+\varepsilon}}. \tag{4.35}$$

Having found a solution $u$ of the parahomological equation (4.19), we determine the constants $\xi, \eta$ in terms of $u$ via the operators $P^x[u], P^y[u]$. We will then be able to cancel all terms but the one linear in $E$ in the paralinearization formula (4.17). Thus the paralinearization formula (4.17) at $u$ reduces to

$$E = \mathbf{Op}^{\mathrm{PM}}\Big(\boldsymbol{B}(E, L[u])M[u]^{-1}\Big)(u-\zeta_0). \tag{4.36}$$

Using Corollary 4.2.1, we estimate

$$\begin{aligned}&\Big\|\mathbf{Op}^{\mathrm{PM}}\Big(\boldsymbol{B}(E, L[u])M[u]^{-1}\Big)(u-\zeta_0)\Big\|_{H^s}\\ &\quad\overset{\text{Prop. 3.3}}{\leq} C_s\big|\boldsymbol{B}(E, L[u])M[u]^{-1}\big|_{L^\infty}\|u-\zeta_0\|_{H^s}\\ &\quad\overset{(4.35)}{\leq} C_s c_2\big(M_Q, |h|_{C^{N_{s+r+3}}}\big)\gamma^{-2}\|e_0\|_{H^{s+2\tau+\varepsilon}}\big\|\boldsymbol{B}(E, L[u])M[u]^{-1}\big\|_{H^{s-(\tau+2)}}\\ &\quad\overset{\text{Cor. 4.2.1}}{\leq} C_s c_2\big(M_Q, |h|_{C^{N_{s+r+3}}}\big)\gamma^{-2}\|e_0\|_{H^{s+2\tau+\varepsilon}}\|E\|_{H^{s-1}}.\end{aligned} \tag{4.37}$$

Note that it is legitimate to apply Corollary 4.2.1, since we already know that $u$ is close to $\zeta_0$, and therefore $E = \mathscr{F}(h_\xi, u) + (0,\mu)^{\mathsf{T}}$ is close to zero. Consequently, if we further require the function $c_0 \ll 1/c_2$, then the factor

$$C_s c_2\big(M_Q, |h|_{C^{N_{s+r+3}}}\big)\gamma^{-3}\|e_0\|_{H^{s+2\tau+\varepsilon}}$$

in the right-hand-side of (4.37) will be $\leq 1/2$ (remembering $e_0$ is of size $O(\gamma^4 c_0)$ and $\gamma$ is assumed to be small). Therefore, (4.37) yields $\|E\|_{H^s} \leq (1/2)\|E\|_{H^{s-1}}$, forcing $E = 0$.

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

To summarize, under the assumption of Theorem 1.1, there is a solution $u$ of the the parahomological equation (4.19) with $\xi = 0$ in (4.19), the constant vector $\mu$ is determined by $u$, and $u$ satisfies

$$E = \mathscr{F}(h, u) + \begin{pmatrix} 0 \\ \mu \end{pmatrix} = 0.$$

Under the assumption of Theorem 1.2, there is a solution $u$ of the the parahomological equation (4.19), the parameters $\xi, \mu$ are determined by $u$, and $u$ satisfies

$$E = \mathscr{F}(h_\xi, u) + \begin{pmatrix} 0 \\ \mu \end{pmatrix} = 0.$$

By Lemma 4.1, the proof of Theorem 1.1–1.2 is complete.

**Remark 4.2.** Similarly as in Subsection 2.4, the right-hand-side of the parainverse equation (4.29) becomes continuously differentiable in $u$ if we assume higher regularity $h \in C^{N_{s+r}+4}$. This is indicated in the general paralinearization theorem 3.5. Thus if the errors $e_0$ and $e_1$ are even smaller, the right-hand-side of (4.29) shall have Lipschitz constant less than 1, and a Banach fixed point argument will apply. In this case, the KAM conjugacy problem falls into the realm of standard implicit function theorems, which will immediately imply, for example, Whitney differentiability with respect to the frequency $\omega$. Compared to what is in the literature, this direct proof is much simpler.

### *4.4. Miscellaneous*

We discuss the generality of the assumption of Theorem 1.1–1.2. The idea of our discussion closely follows [10]. At a first glance, the coverage of Theorem 1.1–1.2 is very limited: the approximate invariant torus around which we seek for a true one has to be the "flat" embedding $\theta \mapsto (\theta, 0)$. However, under suitable geometric transformations, which are elementary and independent from KAM type results, it is possible to reduce quite general situations to this special case.

To be concrete, let the frequency $\omega$ satisfy (1.3). Suppose $h(x, y)$ is an arbitrary Hamiltonian function, not necessarily brought into the form (1.2). Let

$$\zeta(\theta) = \begin{pmatrix} \zeta^x(\theta) \\ \zeta^y(\theta) \end{pmatrix}$$

be an embedding from $\mathbb{T}^n$ to the phase space, not necessarily close to the "flat" embedding, such that $\zeta^x : \mathbb{T}^n \mapsto \mathbb{T}^n$ is a diffeomorphism, and $\zeta(\mathbb{T}^n)$ is an approximately invariant torus with frequency $\omega$ under the flow of $h$: that is,

$$\mathscr{F}(h, \zeta) := X_h(\zeta) - \nabla_\omega \zeta$$

is close to 0. We aim to find suitable symplectic coordinate transformation that reduces this general approximate solution to the special case discussed in Theorem 1.1–1.2.

Recall (4.6): we introduce

$$L[\zeta] = (\partial\zeta)^{\mathsf{T}} \cdot J\partial\zeta \in \boldsymbol{M}_{n\times n}$$

to measure the "lack of isotropy" of the embedding $\zeta$, since the pull-back of the symplectic form $\mathrm{d}x \wedge \mathrm{d}y$ to $\mathbb{T}^n$ via $\zeta$ has exactly the matrix representation $L[\zeta]$. We then have the following, as seen in (4.15):

$$L[\zeta] = \partial\nabla_\omega^{-1}\left[(\partial\zeta)^{\mathsf{T}} \cdot J\mathscr{F}(h, \zeta)\right] - \partial\nabla_\omega^{-1}\left[(\partial\zeta)^{\mathsf{T}} \cdot J\mathscr{F}(h, \zeta)\right]^{\mathsf{T}}. \tag{4.38}$$

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

According to Lemma 6 of [10], if one sets

$$p(\theta) = \Delta^{-1}\partial \cdot L[\zeta] = \left(\Delta^{-1}\sum_{j=1}^{n}\partial_j L_{kj}(\theta)\right)_{1\le k\le n},$$

then the embedded torus

$$\eta(\theta) = \begin{pmatrix}\eta^x(\theta)\\ \eta^y(\theta)\end{pmatrix} := \begin{pmatrix}\zeta^x(\theta)\\ \zeta^y(\theta) - \big(\partial\zeta^x(\theta)\big)^{\mathsf{T}} p(\theta)\end{pmatrix}$$

is isotropic. From (4.38), $p$, hence $\eta - \zeta$, is linear in $\mathscr{F}(h,\zeta)$, and for $s > 1 + n/2$,

$$\|\eta - \zeta\|_{H^s} \le C_s\gamma^{-1}\|\zeta\|_{H^{s+\tau+1}}\|\mathscr{F}(h,\zeta)\|_{H^{s+\tau}},$$

where $C_s$ only depends on $s$. Since

$$\begin{aligned}\mathscr{F}(h,\eta) &= X_h\big(\zeta + (\eta-\zeta)\big) - \nabla_\omega\zeta - \nabla_\omega(\eta-\zeta)\\ &= \mathscr{F}(h,\zeta) + \int_0^1 (DX_h)\big(\zeta + \tau(\eta-\zeta)\big)(\eta-\zeta)\,\mathrm{d}\tau - \nabla_\omega(\eta-\zeta),\end{aligned}$$

it then follows that

$$\|\mathscr{F}(h,\eta)\|_{H^s} \le C_s\gamma^{-1}\big(1 + |h|_{C^{N_{s+r+3}}}\big)\big(1 + \|\zeta\|_{H^{s+\tau+2}}\big)\|\mathscr{F}(h,\zeta)\|_{H^{s+\tau+1}},$$

where $C_s$ only depends on $s$ and $|\omega|$. In other words, if $\zeta(\mathbb{T}^n)$ is an approximately invariant torus for $X_h$, then so is the isotropic torus $\eta(\mathbb{T}^n)$, but the "error of being invariant" is less regular compared to $\zeta(\mathbb{T}^n)$.

Since $\eta$ is an isotropic embedding, the diffeomorphism in phase space

$$G : \begin{pmatrix}x\\ y\end{pmatrix} \mapsto \begin{pmatrix}\eta^x(x)\\ \eta^y(x) + \big(\partial\eta^x(x)\big)^{\mathsf{T}} y\end{pmatrix}$$

is a symplectic one. The isotropic torus $\eta(\mathbb{T}^n)$ is straightened to the "flat" torus $\{y_1 = 0\}$ under the new symplectic coordinate $(x_1, y_1) = G^\iota(x,y)$. If we set

$$H(x_1,y_1) := (h\circ G)(x_1,y_1) = a_0(x_1) + \langle a_1(x_1), y_1\rangle + \frac{1}{2}\langle Q(x_1)y_1, y_1\rangle + O(|y_1|^3) \quad \text{near } y = 0,$$

then the Hamiltonian vector field $X_h$ is changed to $X_H = (DG)^{-1}X_h\circ G$. Furthermore, the "error of being invariant" is transformed to

$$(DG)^{-1}\mathscr{F}(h,\eta) = X_H(\zeta_0) - \begin{pmatrix}\omega\\ 0\end{pmatrix} = \begin{pmatrix}a_1(\theta) - \omega\\ \partial a_0(\theta)\end{pmatrix}.$$

Thus $a_0$ is almost constant, and $a_1$ is almost equal to $\omega$. We then set $S = \nabla_\omega^{-1}(Q - \mathrm{Avg}\,Q)$, and introduce another canonical variable $(x_2, y_2)$, defined by

$$\begin{aligned}x_1 &= x_2 + S(x_1)y_2\\ y_1 &= y_2 - \frac{1}{2}\partial S(x_1)(y_2,y_2).\end{aligned}$$

The symplectic diffeomorphism $\Gamma$ maps the isotropic torus $\{y_2 = 0\}$ to $\{y_1 = 0\}$, while the Hamiltonian function

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

$$H(x_1, y_1) = (a_0(x_1) - \operatorname{Avg} a_0) + \langle a_1(x_1) - \omega, y_1\rangle + \operatorname{Avg} a_0 + \langle \omega, y_2\rangle + \frac{1}{2}\langle (\operatorname{Avg} Q) y_2, y_2\rangle + O(|y_2|^3).$$

We thus find that the first three Taylor coefficients of $H(x_1, y_1)$ with respect to $y_2$ around $y_2 = 0$ will be close to

$$\operatorname{Avg} a_0, \quad \omega, \quad \operatorname{Avg} Q,$$

respectively. Consequently, under the new symplectic coordinate $(x_2, y_2)$, the Hamiltonian function and approximately invariant torus are exactly in the form discussed by Theorem 1.1–1.2.

In conclusion, we have successfully reduced the general situation to the special case of Theorem 1.1 without any assumption on smallness of $\zeta^x, \zeta^y$. For Hamiltonian functions depending on parameters, the argument will be exactly the same.

## A. "Direct method" for conjugacy problems

In this appendix, we explain why we prefer the "indirect method" to study conjugacy problems by comparing it with the "direct method," to be specified shortly. Since this is only a complementary exposition, we will not be concerned with the choice of appropriate function spaces. The somewhat formal computation already suggests that the "direct method" is technically much more cumbersome compared to the "indirect method."

We first sketch how the conjugacy problem for circular maps can be solved in a "direct" way. The idea seems to come from the lecture notes of Hörmander [43]. Define a nonlinear mapping

$$\mathscr{G}(u, \lambda) = \left[\Delta_\alpha u\right] \circ (\mathrm{Id} + u)^\iota + \lambda,$$

which appears in (2.2). Computing the linearization of $\mathscr{G}$ at $(u, \lambda)$ close to $(0, 0)$, we find

$$\begin{aligned} D\mathscr{G}(u, \lambda)(v, \mu) &= \left[\Delta_\alpha v\right] \circ (\mathrm{Id} + u)^\iota \\ &\quad + \left(\left[\Delta_\alpha (u + tv)'\right] \circ (\mathrm{Id} + u + tv)^\iota\right) \cdot \frac{d}{dt}(\mathrm{Id} + u + tv)^\iota\Big|_{t=0} + \mu \\ &= \left[(1 + u' \circ \varrho_\alpha)\Delta_\alpha\left(\frac{v}{1 + u'}\right)\right] \circ (\mathrm{Id} + u)^\iota + \mu. \end{aligned} \tag{A.1}$$

The linearized equation $D\mathscr{G}(u, \lambda)(v, \mu) = h$ is then solved by

$$v = (1 + u')\Delta_\alpha^{-1}\left(\frac{h \circ (\mathrm{Id} + u) - \mu}{1 + u' \circ \varrho_\alpha}\right), \tag{A.2}$$

where $\mu$ is the unique real number making the function inside the bracket to have average 0. Thus, if we consider equation (2.2) instead of (2.1), then *the linearized equation is exactly solvable*.

Note that product and composition of mappings in the grading $\cap_{r\geq 0} C^r_*$ satisfy tame estimates. Using these tame estimates, we find that the solution $(v, \mu)$ of the linearized equation $D\mathscr{G}(u, \lambda)(v, \mu) = h$ satisfies the following tame estimate:

$$|v|_{C^r_*} + |\mu| \lesssim_r \left(1 + |u|_{C^{2+1/2}_*}\right)|h|_{C^{r+s}_*} + |h|_{C^{1/2}_*}\left(1 + |u|_{C^{r+s+1}_*}\right), \quad \text{where } s > \tau + 1.$$

We can finally apply any Nash-Moser type theorem to conclude that, at least for very regular $f$ with small magnitude, there is a solution $u$ of (2.2).

In this case, the paradifferential approach is still valid to produce an alternative proof. We suppose a priori that $u$ is sufficiently smooth. For simplicity of notation, write $W = W(u) = (\mathrm{Id} + u)^\iota$. Let us try

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

to decompose the nonlinear mapping

$$u \mapsto (\Delta_\alpha u) \circ W = (\Delta_\alpha u) \circ (\mathrm{Id} + u)^\iota$$

in the left-hand-side of (2.2) using appropriate paradifferential operators. By Proposition 2.5,

$$(\Delta_\alpha u) \circ W = W^\star \Delta_\alpha u + T_{(\Delta_\alpha u') \circ W} W + \text{smoother remainder}. \tag{A.3}$$

We leave the first term in (A.3) unchanged and concentrate on the paraproduct term, aiming to express $W(u)$ in terms of the paracomposition operator. To do this, we paralinearize the identity $(\mathrm{Id}+u)\circ W(u) = \mathrm{Id}$ to find

$$W(u) + W^\star u + T_{u' \circ W} W(u) = \mathrm{Id} + \text{smoother remainder}.$$

Solving for $W(u)$, we thus obtain

$$W(u) = -T_{(1+u' \circ W)^{-1}} W^\star u + \mathrm{Id} + \text{smoother remainder}. \tag{A.4}$$

We then cite a proposition on conjugation of paradifferential operators using paracomposition, which is a special case of the conjugation theorem proved in [5, 72] or [76]:

**Proposition A.1.** *Let $r > 0, \rho > 1$. Suppose $a(x,\xi) = \sum_{|\alpha| \le m} a_\alpha(x)(i\xi)^\alpha$ is a classical symbol with $C^r_*$ coefficients on $\mathbb{T}^n$. Suppose $\chi : \mathbb{T}^n \mapsto \mathbb{T}^n$ is a $C^\rho_*$ diffeomorphism. Then*

$$\chi^\star T_a = T_{a^\star} \chi^\star + \mathcal{R}_{\mathrm{conj}}(\chi, a),$$

*where*

$$a^*(x,\xi) := \sum_{|\beta|=0}^{[\rho]} \frac{1}{i^{|\beta|}|\beta|!} \partial_y^\beta \partial_\xi^\beta \left( a\big(\chi(x), R(x,y)^{-1}\xi\big) \left| \frac{\det \chi'(y)}{\det R(x,y)} \right| \right) \Big|_{y=x},$$

*with*

$$R(x,y) = \int_0^1 \chi'(tx + (1-t)y)^{\mathrm{T}} dt,$$

*while the remainder operator maps $H^s$ to $H^{s-m+\min(r,\rho-1)}$), and satisfies*

$$\|\mathcal{R}_{\mathrm{conj}}(\chi,a)\|_{\mathcal{L}(H^s, H^{s-m+\min(r,\rho-1)})} \le K_s\big(|\chi|_{C^\rho_*}\big) \sum_{|\alpha|\le m} |a_\alpha|_{C^r_*}.$$

Thus, (A.3) is transformed to

$$\begin{aligned} (\Delta_\alpha u) \circ W &= W^\star \Delta_\alpha u - T_{(\Delta_\alpha u') \circ W} T_{(1+u' \circ W)^{-1}} W^\star u + \text{smoother remainder} \\ &= W^\star \Delta_\alpha u - W^\star T_{(\Delta_\alpha u')/(1+u')} u + \text{smoother remainder} \\ &= W^\star T_{(1+u' \circ \varrho_\alpha)} \Delta_\alpha T_{1/(1+u')} u + \text{smoother remainder}. \end{aligned} \tag{A.5}$$

The original nonlinear equation (2.2) then becomes

$$W^\star T_{(1+u' \circ \varrho_\alpha)} \Delta_\alpha T_{1/(1+u')} u + \lambda = f + \text{smoother remainder}, \tag{A.6}$$

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

which may still be named *parahomological equation*. (A.6) can still be converted to a fixed point type equation:

$$u = T^{-1}_{1/(1+u')} \Delta_\alpha^{-1} T^{-1}_{(1+u')\circ \varrho_\alpha)} (W^\star)^{-1} \big(f - \lambda + \text{smoother remainder}\big). \tag{A.7}$$

The $\lambda$ can be uniquely determined by balancing the mean value, so that $\Delta_\alpha^{-1}$ can be applied.

(A.5) is exactly the *paradifferential version* of the linearization of $\mathscr{G}(u,\lambda)$, or to put simply, the *paralinearization* of $\mathscr{G}(u,\lambda)$. Although quite parallel to the "indirect method," we still prefer the "indirect method" instead. The reason is that the fixed point equation produced by the "indirect method" is technically less involved than the "direct method." Even in this simple illustrative model, the fixed point equation (A.7) requires more paracomposition estimates compared to (2.13): it calls for knowledge about inversion of paracompositions, together with conjugation with paracompositions, while (2.13) relies only on the paralinearization formula.

Let us finally discuss the Hamiltonian conjugacy problem. As pointed out by Féjoz [29], given a Hamiltonian function $h$ close to a normal form, if one aims to find a normal form $H$, a symplectic diffeomorphism $\Gamma$ and a frequency shift $\beta$ solving

$$h = H \circ \Gamma + \beta \cdot y, \tag{A.8}$$

then the linearized equation will be exactly solvable in a neighbourhood of $(\Gamma, \beta, H) = (\mathrm{Id}, 0, h_0)$. On the other hand, if one considers instead

$$h \circ \Gamma - \beta \cdot (y \circ \Gamma) = H,$$

the linearized equation will only admit approximate solution. The form (A.8) then brings a lot of convenience for the Nash-Moser scheme. However, as with the circular map problem discussed earlier, the paralinearization of (A.8) is technically more complicated compared to the "indirect method" since it inevitably involves numbers of paracomposition estimates, though the general idea remains unchanged. In summary, we would prefer the "indirect method" discussed in this paper instead, where we solve the parainverse equation and conclude that it annihilates the nonlinear mapping by a Neumann series argument.

**Acknowledgments.** The authors would like to thank Alain Chenciner, Jacques Féjoz and Mauricio Garay for valuable discussions on classical KAM theory. The authors would also like to thank Yingdu Dong, Giovanni Forni and Nicolò Tedesco for discussions on the computational details.

**Competing interests.** The authors have no competing interests to declare.

## References

[1] T. Alazard, N. Burq and C. Zuily, 'On the water-wave equations with surface tension', *Duke Math. J.* **158**(3) (2011), 413–499.
[2] T. Alazard and G. Métivier, 'Paralinearization of the Dirichlet to Neumann operator, and regularity of three-dimensional water waves', *Comm. Partial Differential Equations* **34**(12) (2009), 1632–1704.
[3] T. Alazard and C. Shao, 'Paracomposition operators and paradifferential reducibility', arXiv preprint arXiv:2410.17211, 2024.
[4] S. Alinhac and P. Gérard, *Opérateurs pseudo-différentiels et théorème de Nash-Moser*, EDP Sciences, 1991.
[5] S. Alinhac, 'Paracomposition et opérateurs paradifférentiels', *Comm. Partial Differential Equations* **11**(1) (1986), 87–121.
[6] V. I. Arnold, 'Proof of a theorem of A. N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations of the Hamiltonian', *Russian Math. Surveys* **18**(5) (1963), 9–36.
[7] V. I. Arnold, 'Small divisor problems in classical and celestial mechanics', *Russian Math. Surveys* **18**(6) (1963), 85–192.
[8] V. I. Arnold, 'Small denominators, I: mappings of the circumference into itself', *Amer. Math. Soc. Transl. Ser.* **2**(46) (1965), 213–288.
[9] J. Bedrossian, N. Masmoudi and C. Mouhot, 'Landau damping: paraproducts and Gevrey regularity', *Ann. PDE* **2** (2016), 1–71.

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

[10] M. Berti and P. Bolle, 'A Nash-Moser approach to KAM theory', in *Hamiltonian Partial Differential Equations and Applications* (Springer, 2015), 255–284.
[11] J.-M. Bony, 'Calcul symbolique et propagation des singularités pour les équations aux dérivées partielles non linéaires', *Ann. Sci. Éc. Norm. Supér.* **14** (1981), 209–246.
[12] J.-B. Bost, 'Tores invariants des systèmes dynamiques hamiltoniens', *Séminaire Bourbaki 1984/85*, Société Mathématique de France, 1986, p. 85.
[13] A. Bounemoura and S. Fischler, 'A Diophantine duality applied to the KAM and Nekhoroshev theorems', *Math. Z.* **275**(3–4) (2013), 1135–1167.
[14] A. Bounemoura and S. Fischler, 'The classical KAM theorem for Hamiltonian systems via rational approximations', *Regul. Chaotic Dyn.* **19** (2014), 251–265.
[15] J. Bourgain, 'Construction of periodic solutions of nonlinear wave equations in higher dimension', *Geom. Funct. Anal.* **5** (1995), 629–639.
[16] J. Bourgain, 'On Melnikov's persistency problem', *Math. Res. Lett.* **4**(4) (1997), 445–458.
[17] J. Bourgain, 'Quasi-periodic solutions of Hamiltonian perturbations of 2D linear Schrödinger equations', *Ann. of Math.* **148**(2) (1998), 363–439.
[18] J. Bricmont, K. Gawedzki, and A. Kupiainen, 'KAM theorem and quantum field theory', *Comm. Math. Phys.* **201** (1999), 699–727.
[19] A. Celletti and L. Chierchia, 'Construction of analytic KAM surfaces and effective stability bounds', *Comm. Math. Phys.* **118**(1) (1988), 119–161.
[20] A. Celletti and L. Chierchia, 'On the stability of realistic three-body problems', *Comm. Math. Phys.* **186**(2) (1997), 413–449.
[21] A. Celletti and L. Chierchia, KAM stability and celestial mechanics, *Mem. Amer. Math. Soc.*, Vol. 187 187.878, American Mathematical Society, 2007.
[22] L. Chierchia and M. Procesi, 'Kolmogorov–Arnold–Moser (KAM) theory for finite and infinite dimensional systems', in *Encyclopedia of Complexity and Systems Science*, Living ed. (Springer, New York, 2018).
[23] W. Craig and C. E. Wayne, 'Newton's method and periodic solutions of nonlinear wave equations', *Comm. Pure Appl. Math.* **46**(11) (1993), 1409–1498.
[24] J.-M. Delort, 'Periodic solutions of nonlinear Schrödinger equations: a paradifferential approach', *Anal. PDE* **4**(5) (2012), 639–676.
[25] D. M. DeTurck, 'Deforming metrics in the direction of their Ricci tensors', *J. Differential Geom.* **18**(1) (1983), 157–162.
[26] L. H. Eliasson, 'Absolutely convergent series expansions for quasi-periodic motions', *Math. Phys. Electron. J.* **4** (1996), 1–33.
[27] A. J. Fang, 'Nonlinear stability of the slowly-rotating Kerr-de Sitter family', arXiv preprint arXiv:2112.07183, 2021.
[28] J. Féjoz, 'Démonstration du "théorème d'Arnold" sur la stabilité du système planétaire (d'après Herman)', *Ergodic Theory Dynam. Systems* **24**(5) (2004), 1521–1582.
[29] J. Féjoz, 'Introduction to KAM theory, with a view to celestial mechanics', in *Variational Methods in Imaging and Geometric Control* (de Gruyter, 2016).
[30] G. Forni, 'Solutions of the cohomological equation for area-preserving flows on compact surfaces of higher genus', *Ann. of Math.* **146**(2) (1997), 295–344.
[31] G. Forni, 'Sobolev regularity of solutions of the cohomological equation', *Ergodic Theory Dynam. Systems* **41**(3) (2021), 685–789.
[32] G. Forni, 'Finite codimension stability of invariant surfaces', arXiv preprint arXiv:2502.00898, 2025.
[33] M. Gage and R. S. Hamilton, 'The heat equation shrinking convex plane curves', *J. Differential Geom.* **23**(1) (1986), 69–96.
[34] L. Grafakos, *Modern Fourier Analysis*, 2nd ed., Grad. Texts Math., vol. 250 (Springer, New York, 2009).
[35] M. Günther, 'On the perturbation problem associated to isometric embeddings of Riemannian manifolds', *Ann. Global Anal. Geom.* **7** (1989), 69–77.
[36] R. S. Hamilton, 'The inverse function theorem of Nash and Moser', *Bull. Amer. Math. Soc. (N.S.)* **7**(1) (1982), 65–222.
[37] R. S. Hamilton, 'Three-manifolds with positive Ricci curvature', *J. Differential Geom.* **17**(2) (1982), 255–306.
[38] M. Hénon, 'Exploration numérique du problème restreint. IV. Masses égales, orbites non périodiques', *Ann. Astrophys.* **3**.2 (1966), 49–66.
[39] M. R. Herman, 'Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations', *Publ. Math. Inst. Hautes Études Sci.* **49** (1979), 5–233.
[40] M. R. Herman, 'Simple proofs of local conjugacy theorems for diffeomorphisms of the circle with almost every rotation number', *Bol. Soc. Brasil. Mat.* **16**(1) (1985), 45–83.
[41] P. Hintz and A. Vasy, 'The global non-linear stability of the Kerr–de Sitter family of black holes', *Acta Math.* **220**(1) (2018), 1–206.
[42] L. Hörmander, 'The boundary problems of physical geodesy', *Arch. Ration. Mech. Anal.* **62**(1) (1976), 1–52.
[43] L. Hörmander, Implicit function theorems, Lectures at Stanford University, summer quarter, 1977.
[44] L. Hörmander, The Nash-Moser theorem and paradifferential operators, in *Analysis, et cetera*, Elsevier, 1990, pp. 429–449.
[45] L. Hörmander, *Lectures on Nonlinear Hyperbolic Differential Equations*, Math. Appl., vol. 26, Springer, 1997.
[46] T. Kato, On nonlinear Schrödinger equations, II. Hs-solutions and unconditional well-posedness, *J. Anal. Math.* 67 (1995), 281–306.

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

[47] Y. Katznelson and D. Ornstein, 'The absolute continuity of the conjugation of certain diffeomorphisms of the circle', *Ergodic Theory Dynam. Systems* **9**(4) (1989), 681–690.
[48] C. E. Kenig, G. Ponce, and L. Vega, 'Well-posedness and scattering results for the generalized Korteweg-de Vries equation via the contraction principle', *Comm. Pure Appl. Math.* **46**(4) (1993), 527–620.
[49] K. Khanin, J. L. Dias, and J. Marklof, 'Multidimensional continued fractions, dynamical renormalization and KAM theory', *Comm. Math. Phys.* **270**(1) (2007), 197–231.
[50] K. M. Khanin and Ya. G. Sinai, 'A new proof of M. Herman's theorem', *Comm. Math. Phys.* **112** (1987), 89–101.
[51] S. Klainerman and G. Ponce, 'Global, small amplitude solutions to nonlinear evolution equations', *Comm. Pure Appl. Math.* **36**(1) (1983), 133–141.
[52] S. Klainerman, 'Global existence for nonlinear wave equations', *Comm. Pure Appl. Math.* **33**(1) (1980), 43–101.
[53] S. Klainerman, 'Long-time behavior of solutions to nonlinear evolution equations', *Arch. Ration. Mech. Anal.* **78**(1) (1982), 73–98.
[54] S. Klainerman, 'Uniform decay estimates and the Lorentz invariance of the classical wave equation', *Comm. Pure Appl. Math.* **38**(3) (1985), 321–332.
[55] A. N. Kolmogorov, 'On the conservation of quasi-periodic motion for a small variation of the Hamiltonian function', *Dokl. Akad. Nauk SSSR* **98** (1954), 527–530.
[56] A. Kriegl and P. W. Michor, *The Convenient Setting of Global Analysis*, Math. Surveys Monogr., vol. 53, American Mathematical Society, 1997.
[57] S. B. Kuksin, 'Hamiltonian perturbations of infinite-dimensional linear systems with an imaginary spectrum', *Funct. Anal. Appl.* **21** (1987), 192–205.
[58] J. Laskar, Michel Hénon and the stability of the solar system, arXiv preprint arXiv:1411.4930, 2014.
[59] R. de la Llave, A. González, À. Jorba, and J. Villanueva, 'KAM theory without action-angle variables', *Nonlinearity* **18**(2) (2005), 855.
[60] R. de la Llave, A tutorial on KAM theory, in *Smooth Ergodic Theory and Its Applications*, Proc. Sympos. Pure Math., vol. 69, American Mathematical Society, 2001, pp. 175–296.
[61] S. Marmi, P. Moussa, and J.-C. Yoccoz, 'The cohomological equation for Roth-type interval exchange maps', *J. Amer. Math. Soc.* **18**(4) (2005), 823–872.
[62] S. Marmi, P. Moussa, and J.-C. Yoccoz, 'Linearization of generalized interval exchange maps', *Ann. of Math.* **176**(3) (2012), 1583–1646.
[63] G. Métivier, *Para-differential Calculus and Applications to the Cauchy Problem for Nonlinear Systems*, *Edizioni della Normale*, 2008.
[64] Y. Meyer, *Remarques sur un théorème de J.-M. Bony*, Université de Paris-Sud, Département de Mathématique, 1980.
[65] Y. Meyer, Régularité des solutions des équations aux dérivées partielles non linéaires, in *Séminaire Bourbaki 1979/80*, Lecture Notes in Math., vol. 842, Springer, 1981, pp. 293–302.
[66] M. Ming and Z. Zhang, 'Well-posedness of the water-wave problem with surface tension', *J. Math. Pures Appl.* **92**(5) (2009), 429–455.
[67] J. Moser, 'On invariant curves of area-preserving mapping of an annulus', *Matematika*, **6**(5) (1962), 51–68.
[68] J. Moser, 'A rapidly convergent iteration method and non-linear partial differential equations—I', *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **20**(2) (1966), 265–315.
[69] J. Moser, 'A rapidly convergent iteration method and non-linear partial differential equations—II', *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **20**(2) (1966), 499–535.
[70] C. Mouhot and C. Villani, 'On Landau damping', *Acta Math.* **207** (2011), 29–201.
[71] J. Nash, 'The imbedding problem for Riemannian manifolds', *Ann. of Math.* **63**(1) (1956), 20–63.
[72] H. Nguyen, 'A sharp Cauchy theory for the 2D gravity-capillary waves', *Ann. Inst. H. Poincaré Anal. Non Linéaire* **34** (2017), 1793–1836.
[73] J. Pöschel, A lecture on the classical KAM theorem, arXiv preprint arXiv:0908.2234, 2009.
[74] J. Pöschel, 'KAM à la R', *Regul. Chaotic Dyn.* **16** (2011), 17–23.
[75] H. Rüssmann, 'KAM iteration with nearly infinitely small steps in dynamical systems of polynomial character', *Discrete Contin. Dyn. Syst. Ser. S* **3**(4) (2010), 683–718.
[76] A. R. Said, 'On paracomposition and change of variables in paradifferential operators', *J. Pseudo-Differ. Oper. Appl.* **14**(2) (2023), 25.
[77] D. Salamon and E. Zehnder, 'KAM theory in configuration space', *Comment. Math. Helv.* **64**(1) (1989), 84–132.
[78] D. Salamon, 'The Kolmogorov-Arnold-Moser theorem', *Math. Phys. Electron. J.* **10**(3) (2004), 1–37.
[79] F. Sergeraert, 'Une généralisation du théorème des fonctions implicites de Nash', *C. R. Acad. Sci. Paris Sér. A* **270** (1970), 861–863.
[80] G. Staffilani, *The Initial Value Problem for Some Dispersive Differential Equations*, Ph.D. thesis, University of Chicago, 1995.
[81] E. Stein and T. Murphy, *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*, Princeton Math. Ser., vol. 3 (Princeton University Press, 1993).
[82] M. Taylor, *Tools for PDE: Pseudodifferential Operators, Paradifferential Operators, and Layer Potentials*, Math. Surveys Monogr., vol. 81 (American Mathematical Society, 2000).

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press

[83] C. E. Wayne, An introduction to KAM theory, preprint, https://math.bu.edu/people/cew/preprints/introkam.pdf, 2008.
[84] J.-C. Yoccoz, 'Conjugaison différentiable des difféomorphismes du cercle dont le nombre de rotation vérifie une condition diophantienne', *Ann. Sci. Éc. Norm. Supér.* **17** (1984), 333–359.
[85] E. Zehnder, 'An implicit function theorem for small divisor problems', *Bull. Amer. Math. Soc.* **80**(1) (1974), 174–179.
[86] E. Zehnder, 'Generalized implicit function theorem with applications to some small divisor problems, I', *Comm. Pure Appl. Math.* **28** (1975), 91–140.
[87] E. Zehnder, 'Generalized implicit function theorem with applications to some small divisor problems, II', *Comm. Pure Appl. Math.* **29** (1976), 49–111.

https://doi.org/10.1017/fms.2026.10207 Published online by Cambridge University Press