ISSN 1560-3547, Regular and Chaotic Dynamics, 2019, Vol. 24, No. 6, pp. 583–606. © Pleiades Publishing, Ltd., 2019.

V. I. Arnold's "Pointwise" KAM Theorem

L. Chierchia* **and** **C. E. Koudjinan****

Dipartimento di Matematica, Università Roma Tre
Largo S. L. Murialdo 1, I-00146 Roma, Italy

Received September 11, 2019; revised October 06, 2019; accepted October 11, 2019

Abstract—We review V.I. Arnold's 1963 celebrated paper [1] *Proof of A. N. Kolmogorov's Theorem on the Conservation of Conditionally Periodic Motions with a Small Variation in the Hamiltonian*, and prove that, optimising Arnold's scheme, one can get "sharp" asymptotic quantitative conditions (as $\varepsilon \to 0$, ε being the strength of the perturbation). All constants involved are explicitly computed.

MSC2010 numbers: 37J40, 37J05, 37J25, 70H08

DOI: 10.1134/S1560354719060017

Keywords: Nearly-integrable Hamiltonian systems, KAM theory, Arnold's Theorem, small divisors, perturbation theory, symplectic transformations

1. INTRODUCTION

a. "*One of the most remarkable of A. N. Kolmogorov's mathematical achievements is his 1954 work on classical mechanics*": this is the beginning of V. I. Arnold's celebrated paper *Proof of A. N. Kolmogorov's Theorem on the Conservation of Conditionally Periodic Motions with a Small Variation in the Hamiltonian* [1], published in 1963, on the occasion of A.N. Kolmogorov's 60th birthday. A few lines after, Arnold adds: "*Its deficiency has been that complete proofs have never been published*".

Even though one could argue whether Kolmogorov's proof in [11] is "complete" or not (see, e. g., [5]), Arnold's paper is certainly a milestone of modern dynamical systems, which not only contains a *complete and detailed* proof of Kolmogorov's theorem, but also introduces new original, technical ideas of enormous impact in finite- and infinite-dimensional systems (for reviews, see, e. g., [2] or [9]).

b. As is well known, Kolmogorov's 1954 theorem in classical mechanics [11] (see also [5]) deals with the persistence, for small ε, of Lagrangian invariant tori of analytic integrable systems governed by a nearly integrable Hamiltonian

$$H(y, x) = K(y) + \varepsilon P(y, x) \tag{1.1}$$

where $(y, x) \in \mathbb{R}^d \times \mathbb{T}^d$ are standard symplectic action-angle variables. In short, the theorem says that:

for small ε, non-degenerate Diophantine unperturbed Lagrangian tori persist

Let us recall that "Diophantine" means that the unperturbed torus $\mathcal{T}_{\omega,0} := \{y_0\} \times \mathbb{T}^d$, which is invariant for the flow ϕ_K^t governed by the integrable Hamiltonian K, is such that the frequency $\omega := K_y(y_0)$ is Diophantine, i. e., it satisfies, for some $\alpha, \tau > 0$,

$$|\omega \cdot k| := \sum_{j=1}^{d} |\omega_j k_j| \geqslant \frac{\alpha}{|k|^\tau}, \qquad \forall\ k \in \mathbb{Z}^d \backslash \{0\}\,; \tag{1.2}$$

*E-mail: luigi@mat.uniroma3.it
**E-mail: ckoudjinan@mat.uniroma3.it

"non-degenerate" means that the Hessian of K at y_0 is invertible; finally, "persists" means that $\mathcal{T}_{\omega,0}$ deforms, for positive small enough ε, into a Lagrangian[1)] torus $\mathcal{T}_{\omega,\varepsilon}$ invariant for ϕ_H^t.

The scheme on which Arnold's proof of Kolmogorov's theorem is based, while sharing two basic ideas of Kolmogorov's approach, namely, the use of a quadratic symplectic iterative method and the idea of keeping fixed the Diophantine frequency of the motion, is quite different from Kolmogorov's scheme in the following respects.

First, for a fixed frequency, Arnold constructs an embedded, Lagrangian invariant torus obtained as a limit of symplectic transformations on action domains shrinking to a single point; in contrast, Kolmogorov conjugates the given Hamiltonian to a complete normal form admitting a Lagrangian invariant torus with the prescribed frequency.

A key difference between these two approaches is that Arnold, at each step of the iteration, needs to control only a finite number of small divisors[2)], which, however, depend on actions (this being the reason for the shrinking to one point of the action domains), while in the denominators appearing in Kolmogorov's scheme there enters only the prefixed Diophantine frequency, allowing one to control at once all small divisors, and also to work with smaller and smaller domains, which contain a fixed open set, allowing one, in the end, to get a genuine symplectic transformation.

A clever quantitative revisitation of Kolmogorov's scheme ([16]) shows that such a scheme leads to optimal asymptotic estimates (as $\varepsilon \to 0$). *We shall show below that this is true also for Arnold's original "pointwise" scheme.*

c. Kolmogorov's and Arnold's schemes are "pointwise" in the sense that they deal with the continuation of a single prefixed unperturbed Lagrangian torus with Diophantine frequency. This is in contrast with versions of the KAM theorem[3)] dealing with the persistence of *sets* of simultaneously persistent invariant tori, see [1, 7, 13, 14]. We point out that, actually, Arnold's original formulation of the KAM theorem in [1] belongs to this second kind of theorem as it states the existence of a set of simultaneously invariant tori, however, *the proof* is pointwise in nature and its scheme is exactly the one we follow closely here. Typically, especially when one is concerned with *lower-dimensional* invariant tori, it is not possible to construct a single torus with some pre-assigned property, but, rather, one obtains "Cantor" families of persistent tori (compare, e. g., [9]).

d. The smallness condition, i. e., how small the perturbation has to be in order for the perturbed invariant torus to exist, depends on local analytic properties of K (and on the analytic norm of P). In particular, the main quantitative "competition" is between ε and the size of the small divisors appearing in the iterative scheme, the size of which may be measured by the "homogeneous Diophantine constant" α (compare Eq. (1.2)) of the prefixed frequency $\omega = K_y(y_0)$.

The most important quantitative relations may be easily understood by looking at explicitly solvable examples, i. e., at integrable systems.

To illustrate this point, let us consider, for example, a simple pendulum with gravity ε,

$$H(y,x) = \frac{1}{2}y^2 + \varepsilon(\cos x - 1)\ , \tag{1.3}$$

1) A Lagrangian manifold is a submanifold of dimension d on which the restriction of the two form $\sum_{j=1}^d dy_j \wedge dx_j$ vanishes.

2) To work with a finite number of divisors, Arnold introduces a Fourier cut-off (depending, in view of analyticity, logarithmically on the size of the perturbation), an idea which has been widely followed also in infinite-dimensional Hamiltonian perturbation theory.

3) Strictly speaking, there does not exists a *KAM theorem* ("KAM" standing for the initials of A. N. Kolmogorv, V. I. Arnold and J. K. Moser), however, normally, it refers to (variations of) Kolmogorov's theorem. Here, we follow this tradition.

viewed as an ε–perturbation of the non-degenerate Hamiltonian $K(y) := \frac{1}{2}y^2$, (here, $d = 1$). The energy zero level $\{H = 0\}$ corresponds to the separatrix, i. e.,

$$y = \pm\sqrt{2\varepsilon(1 - \cos x)}\ ,$$

which shows immediately that in the region $\mathcal{S} := \{|y| \leqslant 2\sqrt{\varepsilon}\}$ there are no homotopically non trivial invariant tori (curves) or, equivalently, no Lagrangian invariant curves, which are graphs over the angle variable ("primary tori"). In other words, the region of action space where unperturbed curves $\{y_0\} \times \mathbb{T}$ may be continued into invariant Lagrangian invariant curves, which stay out of the "singular region" $\mathcal{S}$ are such that:

$$|y_0| > 2\sqrt{\varepsilon}\ . \tag{1.4}$$

Now the resonant relations $|K_y(y_0) \cdot k|$ become, in this one-dimensional example, simply $|y_0||k|$ and the Diophantine condition is, therefore, equivalent to requiring that $\alpha = |y_0|$ (recall (1.2)), and the necessary condition (1.4) becomes

$$\frac{\varepsilon}{\alpha^2} < \frac{1}{4}\ . \tag{1.5}$$

Another fact that can be easily extracted from this example concerns the oscillations of (primary) invariant tori[4)].
For $y_0 > 0$ the invariant (primary) curves are given by

$$y_\varepsilon(x) := \sqrt{y_0^2 + 2\varepsilon(1 - \cos x)} = y_0 + v_\varepsilon(x)\ ,$$

with

$$v_\varepsilon(x) := \frac{2\varepsilon(1 - \cos x)}{y_0 + \sqrt{y_0^2 + 2\varepsilon(1 - \cos x)}}\ .$$

Thus, one has that

$$\operatorname{osc}(y_\varepsilon) = \operatorname{osc}(v_\varepsilon) \geqslant v_\varepsilon(\pi) - v_\varepsilon(0) = \frac{4\varepsilon}{y_0 + \sqrt{y_0^2 + 4\varepsilon}} = \frac{\varepsilon}{y_0}\ \frac{4}{1 + \sqrt{1 + 4\varepsilon/y_0^2}}$$

which, in view of (1.5), yields the relation

$$\operatorname{osc}(v_\varepsilon) \geqslant \frac{4}{1 + \sqrt{2}} \cdot \frac{\varepsilon}{\alpha}. \tag{1.6}$$

Below, we shall prove that the enhanced Arnold scheme leads to a smallness condition of the type (compare (2.4) below)

$$\frac{\varepsilon}{\alpha^2} < c \tag{1.7}$$

(for an ε and α independent constant c), which is in agreement with (1.5).
Furthermore, we shall also show that Arnold's scheme leads to a bound on the oscillations of persistent tori given as graphs $\{y = y_0 + v_*(x), x \in \mathbb{T}^d\}$ of the form (compare (2.6) below)

$$\operatorname{osc}(v_*) \leqslant C \cdot \frac{\varepsilon}{\alpha} \tag{1.8}$$

(for an ε and α independent constant C), which, in view of (1.6), is seen to be optimal (as far as the dependence upon ε and α is concerned), showing the "quantitative sharpness" of Arnold's scheme, on which the proof presented below is based.

Condition (1.7) is also the fundamental quantitative relation needed to evaluate the *measure* of the Kolmogorov set, i. e., the union (in a prefixed bounded domain) of all primary tori.

[4)] A primary Lagrangian torus is a graph over the angles $\{(y, x)|\ y = U(x)\,, x \in \mathbb{T}^d\}$ and its oscillation is given by $\sup_{x,x'} |U(x) - U(x')|$.

Indeed, (1.7) leads to a bound on the Lebesgue measure of the complement of the Kolmogorov set by a constant times $\sqrt{\varepsilon}$ (compare [13, 14]), which again, comparing with the simple pendulum (1.3) — that has a region (the area enclosed by the separatrix) of measure $16\sqrt{\varepsilon}$ free of primary tori — is seen to be asymptotically optimal. It has to be remarked, however, that obtaining such an estimate is quite delicate and far from trivial (for a more detailed discussion on this point, see [3, 8, 12]).

e. As is well known, Arnold's scheme is an iterative Newton scheme yielding a sequence of "renormalised Hamiltonians"

$$H_j := K_j + \varepsilon^{2^j} P_j,$$

so that $H_0 = H$ is the given nearly integrable Hamiltonian (1.1) and, for any j, K_j is integrable (i. e., depends only on the action variable y), real-analytic in a r_j-ball around a point y_j close to y_0 and satisfies:

$$\partial_y K_j(y_j) = \omega := \partial_y K(y_0) \ , \qquad \det \partial_y^2 K_j(y_j) \neq 0 \ , \tag{1.9}$$

which means that at each step the frequency is kept fixed and that the integrable Hamiltonian K_j is non-degenerate. The sequence of Hamiltonians H_j is conjugated, i. e., $H_{j+1} = H_j \circ \phi_j$, with ϕ_j symplectic, closer and closer to the identity. The persistent torus $\mathcal{T}_{\omega,\varepsilon}$ is then obtained as the limit

$$\lim_{j\to+\infty} \phi_0 \circ \cdots \circ \phi_{j-1}(y_j, \mathbb{T}^n) \ .$$

The symplectic transformations ϕ_j's are obtained by solving the classical Hamilton – Jacobi equation so as to remove quadratically the order of the perturbation. In doing this one cannot take into account all small divisors (which are dense) and therefore Arnold introduces a Fourier cut-off κ_j, which allows him to deal with a finite number of small divisors. In view of the exponential decay of Fourier coefficients, κ_j can be taken $\sim \left|\log\left(e^{2^j\|P_j\|}\right)\right|$, which introduces a logarithmic correction[5)], which does not affect the convergence of the scheme. All this is well known.
The problem is to equip the scheme with "optimal" quantitative estimates, which may lead, in the end, to the above sharp asymptotic bounds. This involves careful choices of various parameters entering the scheme (see § 3.2) and, in particular, *it is crucial to treat the first step in a different way with respect to the remaining steps*: this technical, but important, aspect is explained in Remark 1 below.

f. V. I. Arnold pointed out that his proof extended with little changes to the *iso-energetically non-degenerate* case, i. e., when the energy is prescribed and the unperturbed Hamiltonian satisfies the condition[6)]

$$\det \begin{pmatrix} \partial_y^2 K & \partial_y K \\ \partial_y K & 0 \end{pmatrix}\Big|_{y=y_0} \neq 0\,. \tag{1.10}$$

Indeed, it would not be difficult to adapt our improved Arnold's scheme also to the iso-energetically non-degenerate case, proving the sharpness of the asymptotic smallness conditions also in this case.

g. Finally, we mention that the quantitative estimates provided in this paper could be used to improve the (exponentially long) stability time of "nearly-invariant tori", introduced in [10].

[5)] For full details, see § 3.1 below and, in particular, "Step 1: Construction of Arnold's transformation".

[6)] The matrix in (1.10) is a $(d+1)\times(d+1)$ matrix, where the upper right corner $\partial_y K$ has to be interpreted as a column vector, while the lower left corner is a raw vector and the zero is a scalar. The condition expresses the fact that the map $(y,\lambda)\mapsto(\lambda\partial_y K, K)$ is locally invertible.

2. NOTATION AND QUANTITATIVE STATEMENT OF ARNOLD'S THEOREM

- For $d \in \mathbb{N} := \{1, 2, 3, \ldots\}$ and $x, y \in \mathbb{C}^d$, we let $x \cdot y := x_1 \bar{y}_1 + \cdots + x_d \bar{y}_d$ be the standard inner product; $|x|_1 := \sum_{j=1}^{d} |x_j|$ be the 1-norm, and $|x| := \max_{1 \leqslant j \leqslant n} |x_j|$ be the sup-norm.

- $\mathbb{T}^d := \mathbb{R}^d / 2\pi\mathbb{Z}^d$ is the standard d-dimensional (flat) torus.

- $\pi_1 \colon \mathbb{C}^d \times \mathbb{C}^d \ni (y, x) \longmapsto y$ and $\pi_2 \colon \mathbb{C}^d \times \mathbb{C}^d \ni (y, x) \longmapsto x$ are the projections on the first and second component, respectively.

- For $\alpha > 0$, $\tau \geqslant d - 1 \geqslant 1$,

$$\Delta_\alpha^\tau := \left\{ \omega \in \mathbb{R}^d : |\omega \cdot k| \geqslant \frac{\alpha}{|k|_1^\tau}, \quad \forall\, 0 \neq k \in \mathbb{Z}^d \right\} \tag{2.1}$$

  is the set of (α, τ)-Diophantine numbers in $\mathbb{R}^d$.

- For $r, s > 0$, $y_0 \in \mathbb{C}^d$, we denote:

$$\begin{aligned} \mathbb{T}_s^d &:= \left\{ x \in \mathbb{C}^d : |\operatorname{Im} x| < s \right\} / 2\pi\mathbb{Z}^d, \\ B_r(y_0) &:= \left\{ y \in \mathbb{R}^d : |y - y_0| < r \right\}, \qquad (y_0 \in \mathbb{R}^d), \\ D_r(y_0) &:= \left\{ y \in \mathbb{C}^d : |y - y_0| < r \right\}, \quad D_{r,s}(y_0) := D_r(y_0) \times \mathbb{T}_s^d. \end{aligned}$$

- If $\mathbb{1}_d := \operatorname{diag}(1)$ is the unit $(d \times d)$ matrix, we denote the standard symplectic matrix by

$$\mathbb{J} := \begin{pmatrix} 0 & -\mathbb{1}_d \\ \mathbb{1}_d & 0 \end{pmatrix}.$$

- For $y_0 \in \mathbb{R}^d$, $\mathcal{A}_{r,s}(y_0)$ denotes the Banach space of real-analytic functions with bounded holomorphic extensions to $D_{r,s}(y_0)$, with norm

$$\|\cdot\|_{r,s,y_0} := \sup_{D_{r,s}(y_0)} |\cdot| \,.$$

  We also denote:

$$\|\cdot\|_{r,y_0} := \sup_{D_r(y_0)} |\cdot| \,, \qquad \|\cdot\|_s := \sup_{\mathbb{T}_s^d} |\cdot| \,.$$

- We equip $\mathbb{C}^d \times \mathbb{C}^d$ with the canonical symplectic form

$$\varpi := dy \wedge dx = dy_1 \wedge dx_1 + \cdots + dy_d \wedge dx_d$$

  and denote by ϕ_H^t the associated Hamiltonian flow governed by the Hamiltonian $H(y, x)$, $y, x \in \mathbb{C}^d$, i. e., $z(t) := \phi_H^t(y, x)$ is the solution of the Cauchy problem $\dot{z} = \mathbb{J}\nabla H(z)$, $z(0) = (y, x)$.

- Given a linear operator $\mathcal{L}$ from the normed space $(V_1, \|\cdot\|_1)$ into the normed space $(V_2, \|\cdot\|_2)$, its "operator-norm" is given by

$$\|\mathcal{L}\| := \sup_{x \in V_1 \setminus \{0\}} \frac{\|\mathcal{L}x\|_2}{\|x\|_1}, \quad \text{so that} \quad \|\mathcal{L}x\|_2 \leqslant \|\mathcal{L}\|\,\|x\|_1 \quad \text{for any} \quad x \in V_1.$$

- Given $\omega \in \mathbb{R}^d$, the directional derivative of a C^1 function f with respect to ω is given by

$$D_\omega f := \omega \cdot f_x = \sum_{j=1}^{d} \omega_j f_{x_j}.$$

- If f is a (smooth or analytic) function on $\mathbb{T}^d$, its Fourier expansion is given by

$$f = \sum_{k\in\mathbb{Z}^d} f_k\, e^{ik\cdot x}, \qquad f_k := \frac{1}{(2\pi)^d}\int_{\mathbb{T}^d} f(x)\, e^{-ik\cdot x}\, dx\,,$$

(where, as usual, $e := \exp(1)$ denotes the Neper number and i the imaginary unit). We also set:

$$\langle f\rangle := f_0 = \frac{1}{(2\pi)^d}\int_{\mathbb{T}^d} f(x)\, dx\,, \qquad (\mathbf{p}_N f)(x) := \sum_{|k|_1\leqslant N} f_k\, e^{ik\cdot x},\ N>0\,,$$

$\mathbf{p}_N$ being the Fourier projection onto the Fourier modes with $|k|_1 \leqslant N$; notice that $\langle\cdot\rangle = \mathbf{p}_0(\cdot)$.

We are ready to formulate a quantitative version of Arnold's theorem[7)].

Theorem A. *Let $d \geqslant 2$; $\tau \geqslant d-1$; $\alpha, r, \varepsilon > 0$; $0 < s_* < s \leqslant 1$; $y_0 \in \mathbb{R}^d$; $K, P \in \mathcal{A}_{r,s}(y_0)$; $H := K + \varepsilon P$. Assume that*

$$\begin{cases} \omega := \partial_y K(y_0) \in \Delta^\tau_\alpha\,, \\ \det(\partial_y^2 K(y_0)) \neq 0\,. \end{cases} \tag{2.2}$$

Define:

$$T := \partial_y^2 K(y_0)^{-1},\ \mathsf{P} := \|P\|_{r,s,y_0},\ \mathsf{K} := \|\partial_y^2 K\|_{r,y_0},\ \mathsf{T} := \|T\|\,,\ \theta := \mathsf{T}\mathsf{K},$$

and denote by ϵ the rescaled smallness parameter:

$$\boxed{\epsilon := \mathsf{K}\mathsf{P}\frac{\varepsilon}{\alpha^2}.} \tag{2.3}$$

There exist constants $1 < \mathsf{C} < \mathsf{C}_*$ depending only on d and τ, such that, if $a := 6\tau + 3d + 8$ and

$$\alpha \leqslant \frac{r}{\mathsf{T}} \qquad \textit{and} \qquad \epsilon \leqslant \epsilon_* := \frac{(s-s_*)^a}{\mathsf{C}_*\,\theta^4}\,, \tag{2.4}$$

then there exists a real–analytic embedding

$$\phi_*\colon x \in \mathbb{T}^d_{s_*} \mapsto \phi_*(x) := \phi_{\mathrm{e}}(y_0, x) + \big(v_*(x), u_*(x)\big) \in D_{r,s}(y_0),$$

where ϕ_{e} is the trivial embedding

$$\phi_{\mathrm{e}}\colon x \in \mathbb{T}^d \to (y_0, x),$$

such that the d–torus

$$\mathcal{T}_{\omega,\varepsilon} := \phi_*\left(\mathbb{T}^d\right) \tag{2.5}$$

is a Lagrangian torus satisfying

$$\phi_H^t \circ \phi_*(x) = \phi_*(x + \omega t), \quad \forall\ x \in \mathbb{T}^d_{s_*}\ \ \forall\ t \in \mathbb{R}\,.$$

---

7) To avoid introducing too many symbols, we use capital straight style for positive constants ($\mathsf{P}, \mathsf{K}, \mathsf{T}, \mathsf{C}, \ldots$), while usually capital normal style is used for functions or matrices ($K, P, H, T, \ldots$).

Furthermore,

$$\max\left\{\|u_*\|_{s_*},\, \frac{1}{2\,e}\,\|\partial_x u_*\|_{s_*},\, \frac{\mathsf{K}}{\alpha}\,\|v_*\|_{s_*}\right\} \leqslant \frac{\mathsf{C}\;\theta^3}{(s-s_*)^a}\;\epsilon \leqslant \frac{1}{4\,e}. \tag{2.6}$$

Remarks and addenda.

(i) θ is a measure of the local "torsion" and is a number greater than or equal to one:

$$\theta := \mathsf{TK} \geqslant \mathsf{T}\|K_{yy}(y_0)\| \geqslant \|T\|\|K_{yy}(y_0)\| = \|T\|\|T^{-1}\| \geqslant 1\ . \tag{2.7}$$

(ii) Notice that the estimate on v_* in (2.6) implies that the maximal action oscillation of the torus $\mathcal{T}_{\omega,\varepsilon}$ is bounded by a constant times $\alpha\epsilon$, which in view of (2.3) is $\sim \varepsilon/\alpha$ as advertised in (1.8).

(iii) All numerical constants are explicitly "computed" during the proof. A complete list of them, including the definitions of C_* and C, is given in Appendix A.

(iv) The torus $\mathcal{T}_{\omega,\varepsilon}$ is *Kolmogorov non-degenerate*. More precisely, H can be put in Kolmogorov's normal form with non-degenerate quadratic part: there exists a symplectic transformation ϕ close to ϕ_{e}, for which

$$H\circ\phi(y,x) = E + \omega\cdot y + Q(y,x) \quad \text{such that} \quad \det\langle Q_{yy}(0,\cdot)\rangle \neq 0;$$

for details, see Appendix B.

(v) The value of ϵ_* in (2.4) is not optimal. In Remark 2 a better (still not optimal) value is given.

(vi) The dependence of the invariant torus $\mathcal{T}_{\omega,\varepsilon}$ on ε is analytic. More generally, if $H = H(y,x;z)$ is real-analytic also in $z\in V$, V being some open set in $\mathbb{C}^m$, and all the above norms are uniform in $z\in V$, then the invariant torus $\mathcal{T}_{\omega,z}$ is real-analytic in V. This is an obvious corollary of Weierstrass's theorem on uniform limits of holomorphic functions, in view of the uniformity of the limits in the proof.

3. PROOF

3.1. Arnold's Scheme: the Basic Step

The next Lemma describes Arnold's basic KAM step, on which Arnold's scheme is based. Its quantitative formulation involves a few constants, which are defined as follows:

$$\nu := \tau+1\,,\quad \mathsf{C}_0 := 4\sqrt{2}\left(\frac{3}{2}\right)^{2\nu+d}\int_{\mathbb{R}^d}\left(|y|_1^\nu + |y|_1^{2\nu}\right)\,e^{-|y|_1}dy\,,$$
$$\mathsf{C}_1 := 2\left(\frac{3}{2}\right)^{\nu+d}\int_{\mathbb{R}^d}|y|_1^\nu\,e^{-|y|_1}dy\,,$$
$$\mathsf{C}_2 := 2^{3d}d\,,\quad \mathsf{C}_3 := \left(d^2\mathsf{C}_1^2 + 6d\mathsf{C}_1 + \mathsf{C}_2\right)\sqrt{2}\,,\quad \mathsf{C}_4 := \max\left\{6d^2\mathsf{C}_0,\, \mathsf{C}_3\right\}.$$

Lemma 1. *Let*[8] $r>0$, $0<2\sigma<s\leqslant 1$, $\mathsf{y}\in\mathbb{R}^d$, $K,P\in\mathcal{A}_{r,s}(\mathsf{y})$ *and consider the Hamiltonian parametrised by* $\varepsilon>0$

$$H(y,x;\varepsilon) := K(y) + \varepsilon P(y,x).$$

Assume that

$$\det K_{yy}(\mathsf{y}) \neq 0\,,\qquad \omega := K_y(\mathsf{y}) \in \Delta_\alpha^\tau\,,$$

[8] K and P stand here for generic real-analytic Hamiltonians which later on will, respectively, play the roles of K_j and P_j, and y, r, the roles of y_j, r_j in the iterative step.

and let K, T *and* P *be positive numbers such that*

$$\|K_{yy}\|_{r,\mathsf{y}} \leqslant \mathsf{K}, \qquad \|T\| \leqslant \mathsf{T}, \qquad \|P\|_{r,s,\mathsf{y}} \leqslant \mathsf{P}, \tag{3.1}$$

where $T := K_{yy}(\mathsf{y})^{-1}$.
Now let $\lambda, \check{r}, \bar{r}$ *be positive numbers such that:*

$$\lambda \geqslant \log\Big(\sigma^{2\nu+d}\frac{\alpha^2}{\varepsilon\,\mathsf{PK}}\Big), \quad \check{r} \leqslant \frac{5}{24d}\frac{r}{\mathsf{TK}}\ , \quad \bar{r} \leqslant \min\Big\{\check{r},\ \frac{\alpha}{2d\mathsf{K}\kappa^{\tau+1}}\Big\}\ , \tag{3.2}$$

where

$$\kappa := \frac{4\lambda}{\sigma}\ .$$

Finally, define

$$\mathsf{L} := \mathsf{P}\max\Big\{\frac{40d\mathsf{T}^2\mathsf{K}}{r^2}\sigma^{-(\nu+d)},\ \frac{\mathsf{C}_4}{\sqrt{2}}\max\big\{1, \frac{\alpha}{r\mathsf{K}}\big\}\frac{\mathsf{K}}{\alpha^2}\sigma^{-2(\nu+d)}\Big\}\ , \quad \bar{s} := s - \frac{2}{3}\sigma, \quad s' := s-\sigma\ .$$

Then, if

$$\varepsilon\,\mathsf{L} \leqslant \frac{\sigma}{3}, \tag{3.3}$$

there exist $\mathsf{y}' \in \mathbb{R}^d$ *and a symplectic change of coordinates*

$$\phi' = \mathrm{id} + \varepsilon\tilde{\phi}\colon D_{\bar{r}/2,s'}(\mathsf{y}') \to D_{2r/3,\bar{s}}(\mathsf{y}), \tag{3.4}$$

such that

$$\begin{cases} H\circ\phi' =: H' =: K' + \varepsilon^2 P'\ , \\ \partial_{y'}K'(\mathsf{y}') = \omega, \quad \det\partial^2_{y'}K'(\mathsf{y}') \neq 0, \end{cases} \tag{3.5}$$

where

$$K' := K + \varepsilon\widetilde{K} := K + \varepsilon\big\langle P(y',\cdot)\big\rangle.$$

Moreover, letting

$$\big(\partial^2_{y'}K'(\mathsf{y}')\big)^{-1} =: T + \varepsilon\,\widetilde{T},$$

the following estimates hold:

$$\|\partial^2_{y'}\widetilde{K}\|_{r/2,\mathsf{y}} \leqslant \mathsf{KL}\,, \quad |\mathsf{y}'-\mathsf{y}| \leqslant \frac{8\varepsilon\,\mathsf{TP}}{r}, \quad \|\widetilde{T}\| \leqslant \mathsf{TL}\,, \tag{3.6}$$

$$\max\{\|\partial_x\pi_2\tilde{\phi}\|_{s'}, \|\mathsf{W}\,\tilde{\phi}\|_{\bar{r}/2,s',\mathsf{y}'}\} \leqslant d^{-2}\sigma^{d-1}\mathsf{L}, \quad \|P'\|_{\bar{r}/2,s',\mathsf{y}'} \leqslant \mathsf{LP}, \tag{3.7}$$

where

$$\mathsf{W} := \begin{pmatrix} \max\{\frac{\mathsf{K}}{\alpha}\,, \frac{1}{r}\}\ \mathbb{1}_d & 0 \\ 0 & \mathbb{1}_d \end{pmatrix}.$$

Observe that

$$\sigma^{-2(\nu+d)}\varepsilon\ \mathsf{PK}/\alpha^2 \leqslant (\sqrt{2}/\mathsf{C}_4)\ \varepsilon\ \mathsf{L},$$

so that (3.3) implies

$$\frac{\varepsilon\,\mathsf{PK}}{\alpha^2} < \frac{\sigma^{2\nu+d}}{e}\ ,$$

which, in particular, implies that $\lambda > 1$ and $\kappa > 4$.

Proof.

Step 1: Construction of Arnold's transformation

We seek a near-identity symplectic transformation

$$\phi' \colon D_{r_1,s_1}(\mathsf{y}') \to D_{r,s}(\mathsf{y}),$$

with $D_{r_1,s_1}(\mathsf{y}') \subset D_{r,s}(\mathsf{y})$, generated by a generating function[9)] of the form $y' \cdot x + \varepsilon g(y', x)$, so that

$$\phi' \colon \begin{cases} y = y' + \varepsilon g_x(y', x) \\ x' = x + \varepsilon g_{y'}(y', x), \end{cases} \tag{3.8}$$

such that

$$\begin{cases} H' \coloneqq H \circ \phi' = K' + \varepsilon^2 P' \ , \\ \partial_{y'} K'(\mathsf{y}') = \omega, \quad \det \partial^2_{y'} K'(\mathsf{y}') \neq 0. \end{cases} \tag{3.9}$$

By Taylor's formula, we get[10)]

$$\begin{aligned} H(y' + \varepsilon g_x(y', x), x) &= K(y') + \varepsilon \widetilde{K}(y') + \varepsilon \left[K'(y') \cdot g_x + \mathbf{p}_\kappa P(y', \cdot) - \widetilde{K}(y') \right] \\ &\quad + \varepsilon^2 \left(P^{(1)} + P^{(2)} + P^{(3)} \right) (y', x) \\ &= K'(y') + \varepsilon \left[K'(y') \cdot g_x + \mathbf{p}_\kappa P(y', \cdot) - \widetilde{K}(y') \right] + \varepsilon^2 P_+(y', x), \end{aligned} \tag{3.10}$$

with $\kappa > 0$, which will be chosen large enough so that $P^{(3)} = O(\varepsilon)$ and

$$\begin{cases} P_+ \ \coloneqq P^{(1)} + P^{(2)} + P^{(3)} \\ P^{(1)} \coloneqq \dfrac{1}{\varepsilon^2} \left[K(y' + \varepsilon g_x) - K(y') - \varepsilon K_y(y') \cdot g_x \right] = \displaystyle\int_0^1 (1-t) K_{yy}(\varepsilon t g_x) \cdot g_x \cdot g_x dt \\ P^{(2)} \coloneqq \dfrac{1}{\varepsilon} \left[P(y' + \varepsilon g_x, x) - P(y', x) \right] = \displaystyle\int_0^1 P_y(y' + \varepsilon t g_x, x) \cdot g_x dt \\ P^{(3)} \coloneqq \dfrac{1}{\varepsilon} \left[P(y', x) - \mathbf{p}_\kappa P(y', \cdot) \right] = \dfrac{1}{\varepsilon} \displaystyle\sum_{|n|_1 > \kappa} P_n(y')\, e^{in \cdot x}. \end{cases} \tag{3.11}$$

By the non-degeneracy condition $\det K_{yy}(\mathsf{y}) \neq 0$, for ε small enough (to be made precise below), $\det \partial^2_{y'} K'(\mathsf{y}) \neq 0$ and, therefore, by the standard inverse function theorem (see, e. g., Lemma 5), there exists a unique $\mathsf{y}' \in D_r(\mathsf{y})$ such that the second part of (3.9) holds. In view of (3.10), in order to get the first part of (3.9), we need to find g such that $K_y(y') \cdot g_x + \mathbf{p}_\kappa P(y', \cdot) - \widetilde{K}(y')$ vanishes; such a g is indeed given by

$$g \coloneqq \sum_{0 < |n|_1 \leqslant \kappa} \frac{-P_n(y')}{i K_y(y') \cdot n}\, e^{in \cdot x}, \tag{3.12}$$

provided that

$$K_y(y') \cdot n \neq 0, \quad \forall\, 0 < |n|_1 \leqslant \kappa, \quad \forall\, y' \in D_{r_1}(\mathsf{y}') \quad (\subset D_r(\mathsf{y}))\,. \tag{3.13}$$

But, in fact, since $K_y(\mathsf{y})$ is rationally independent, then, given any $\kappa > 0$, there exists $\bar{r} \leqslant r$ such that

$$K_y(y') \cdot n \neq 0, \quad \forall\, 0 < |n|_1 \leqslant \kappa, \quad \forall\, y' \in D_{\bar{r}}(\mathsf{y}). \tag{3.14}$$

---

9) Following the classical approach of Arnold, we use generating functions to construct symplectic transformations. Of course, one could also use the equivalent method of time-one Hamiltonian flows (or Lie series).

10) Recall (§2) that $\langle \cdot \rangle$ stands for the average over $\mathbb{T}^d$ and that $\mathbf{p}_N$ is the Fourier projection onto modes with $|k|_1 \leqslant N$.

The last step is to invert the function $x \mapsto x + \varepsilon g_{y'}(y', x)$ in order to define P'. By the inverse function theorem, for ε small enough, the map $x \mapsto x + \varepsilon g_{y'}(y', x)$ admits a real-analytic inverse of the form

$$\varphi_\varepsilon(y', x') = x' + \varepsilon \widetilde{\varphi}_\varepsilon(y', x'), \tag{3.15}$$

so that Arnold's symplectic transformation is given by

$$\phi' \colon (y', x') \mapsto \begin{cases} y = y' + \varepsilon g_x(y', \varphi_\varepsilon(y', x')) \\ x = \varphi_\varepsilon(y', x') = x' + \varepsilon \widetilde{\varphi}_\varepsilon(y', x'). \end{cases} \tag{3.16}$$

Hence, (3.9) holds with

$$P'(y', x') \coloneqq P_+\big(y', \varphi_\varepsilon(y', x')\big). \tag{3.17}$$

Step 2: Quantitative estimates

First of all, notice that from the definitions of $\bar{r}$ and $\check{r}$ it follows that

$$\bar{r} \leqslant \check{r} \leqslant \frac{5r}{24d} < \frac{r}{2}\,. \tag{3.18}$$

We begin by extending the "Diophantine condition w.r.t. K_y" uniformly to $D_{\bar{r}}(\mathsf{y})$ up to order κ. Indeed, by the mean value inequality and $K_y(\mathsf{y}) = \omega \in \Delta^\tau_\alpha$, we get, for any $0 < |n|_1 \leqslant \kappa$ and any $y' \in D_{\bar{r}}(\mathsf{y})$,

$$\begin{aligned} |K_y(y') \cdot n| &= |\omega \cdot n + (K_y(y') - K_y(\mathsf{y})) \cdot n| \geqslant q|\omega \cdot n| \left(1 - d\frac{\|K_{yy}\|_{\bar{r},\mathsf{y}}}{|\omega \cdot n|}|n|_1 \bar{r}\right) \\ &\geqslant q\frac{\alpha}{|n|_1^\tau}\left(1 - \frac{d\mathsf{K}}{\alpha}|n|_1^{\tau+1}\bar{r}\right) \geqslant q\frac{\alpha}{|n|_1^\tau}\left(1 - \frac{d\mathsf{K}}{\alpha}\kappa^{\tau+1}\bar{r}\right) \geqslant \frac{\alpha}{2|n|_1^\tau}, \end{aligned} \tag{3.19}$$

so that, by Fourier estimates (Lemma 4**(ii)**), we have

$$\begin{aligned} \|g_x\|_{\bar{r},\bar{s},\mathsf{y}} &\overset{\text{def}}{=} \sup_{D_{\bar{r},\bar{s}}(\mathsf{y})} \left| \sum_{0<|n|_1\leqslant\kappa} \frac{nP_n(y')}{K_y(y')\cdot n} e^{in\cdot x} \right| \leqslant \sum_{0<|n|_1\leqslant\kappa} \frac{\|P_n\|_{\bar{r},\bar{s},\mathsf{y}}}{|K_y(y')\cdot n|}|n|_1 e^{\left(s-\frac{2}{3}\sigma\right)|n|_1} \\ &\leqslant \sum_{0<|n|_1\leqslant\kappa} \mathsf{P}\, e^{-s|n|_1}\frac{2|n|_1^\nu}{\alpha} e^{\left(s-\frac{2}{3}\sigma\right)|n|_1} \leqslant \frac{2\mathsf{P}}{\alpha}\sum_{n\in\mathbb{Z}^d}|n|_1^\nu e^{-\frac{2}{3}\sigma|n|_1} \\ &\leqslant \frac{2\mathsf{P}}{\alpha}\int_{\mathbb{R}^d}|y|_1^\nu e^{-\frac{2}{3}\sigma|y|_1}dy = \left(\frac{3}{2\sigma}\right)^{\nu+d}\frac{2\mathsf{P}}{\alpha}\int_{\mathbb{R}^d}|y|_1^\nu e^{-|y|_1}dy = \mathsf{C}_1\frac{\mathsf{P}}{\alpha}\sigma^{-(\nu+d)}, \end{aligned}$$

$$\begin{aligned} \|\partial_{y'} g\|_{\bar{r},\bar{s},\mathsf{y}} &\overset{\text{def}}{=} \sup_{D_{\bar{r},\bar{s}}(\mathsf{y})} \left| \sum_{0<|n|_1\leqslant\kappa} \left(\frac{\partial_y P_n(y')}{K_y(y')\cdot n} - P_n(y')\frac{K_{yy}(y')n}{(K_y(y')\cdot n)^2}\right) e^{in\cdot x} \right| \\ &\leqslant \sum_{0<|n|_1\leqslant\kappa} \sup_{D_{\bar{r}}(\mathsf{y})} \left(\frac{\|(P_y)_n\|_{\bar{r},s,\mathsf{y}}}{|K_y(y')\cdot n|} + \|P_n\|_{r,s,\mathsf{y}}\frac{\|K_{yy}\|_{r,\mathsf{y}}|n|_1}{|K_y(y')\cdot n|^2}\right) e^{\left(s-\frac{2}{3}\sigma\right)|n|_1} \\ &\overset{(3.1)+(3.19)}{\leqslant} \sum_{0<|n|_1\leqslant\kappa} \left(\frac{\mathsf{P}}{r-\bar{r}} e^{-s|n|_1}\frac{2|n|_1^\tau}{\alpha} + \mathsf{P}\, e^{-s|n|_1}\mathsf{K}|n|_1\left(\frac{2|n|_1^\tau}{\alpha}\right)^2\right) e^{\left(s-\frac{2}{3}\sigma\right)|n|_1} \\ &\overset{(3.18)}{\leqslant} \frac{4\mathsf{P}}{\alpha^2 r}\sum_{0<|n|_1\leqslant\kappa}\left(|n|_1^\tau\alpha + r\mathsf{K}|n|_1^{2\tau+1}\right) e^{-\frac{2}{3}\sigma|n|_1} \\ &\leqslant \max\{\alpha, r\mathsf{K}\}\frac{4\mathsf{P}}{\alpha^2 r}\sum_{0<|n|_1\leqslant\kappa}\left(|n|_1^\tau + |n|_1^{2\tau+1}\right) e^{-\frac{2}{3}\sigma|n|_1} \\ &\leqslant \max\left\{1, \frac{\alpha}{r\mathsf{K}}\right\}\frac{4\mathsf{P}\mathsf{K}}{\alpha^2}\int_{\mathbb{R}^d}\left(|y|_1^\tau + |y|_1^{2\tau+1}\right) e^{-\frac{2}{3}\sigma|y|_1}dy \end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{3}{2\sigma}\right)^{2\tau+d+1} \max\left\{1, \frac{\alpha}{r\mathsf{K}}\right\} \frac{4\mathsf{PK}}{\alpha^2} \int_{\mathbb{R}^d} \left(|y|_1^\tau + |y|_1^{2\tau+1}\right) e^{-|y|_1} dy \\
&\leqslant \frac{\mathsf{C}_0}{\sqrt{2}} \max\left\{1, \frac{\alpha}{r\mathsf{K}}\right\} \frac{\mathsf{PK}}{\alpha^2} \sigma^{-(2\tau+d+1)} < \overline{\mathsf{L}},
\end{aligned}
$$

where

$$
\overline{\mathsf{L}} \coloneqq 6\ \frac{\mathsf{C}_0}{\sqrt{2}} \max\left\{1, \frac{\alpha}{r\mathsf{K}}\right\} \frac{\mathsf{PK}}{\alpha^2} \sigma^{-(2\nu+d+1)}\ .
$$

Analogously,

$$
\|\partial^2_{y'x} g\|_{\bar{r},\bar{s},\mathsf{y}} \leqslant \frac{\mathsf{C}_0}{\sqrt{2}} \max\left\{1, \frac{\alpha}{r\mathsf{K}}\right\} \frac{\mathsf{PK}}{\alpha^2} \sigma^{-(2\nu+d)} \leqslant \overline{\mathsf{L}},
$$

and, by Cauchy's estimate (Lemma 4(**i**)), we get

$$
\|\partial^3_{y'xx} g\|_{\bar{r},s'',\mathsf{y}} \leqslant \frac{6\mathsf{C}_0}{\sqrt{2}} \max\left\{1, \frac{\alpha}{r\mathsf{K}}\right\} \frac{\mathsf{PK}}{\alpha^2} \sigma^{-(2\nu+d+1)} = \overline{\mathsf{L}}, \tag{3.20}
$$

where

$$
s'' \coloneqq s - \frac{5}{6}\sigma \quad \text{and} \quad \|\partial^3_{y'xx} g\|_{\bar{r},s'',\mathsf{y}} \coloneqq \sup_{D_{\bar{r},s''}(\mathsf{y})} \max\{|\partial^3_{y'_i x_j x_k} g| \ : \ i,j,k = 1,\cdots,d\}.
$$

Also,

$$
\|\widetilde{K}_y\|_{r/2,\mathsf{y}} = \|\langle P_y\rangle\|_{r/2,\mathsf{y}} \leqslant \|P_y\|_{r/2,\bar{s},\mathsf{y}} \leqslant \frac{\mathsf{P}}{r-\frac{r}{2}} \leqslant \frac{2\mathsf{P}}{r}\ ,
$$

$$
\|\partial^2_{y'} \widetilde{K}\|_{r/2,\mathsf{y}} = \|\langle P_{yy}\rangle\|_{r/2,\mathsf{y}} \leqslant \|P_{yy}\|_{r/2,\bar{s},\mathsf{y}} \leqslant \frac{\mathsf{P}}{(r-\frac{r}{2})^2} \leqslant \frac{4\mathsf{P}}{r^2} \leqslant \mathsf{KL}\ .
$$

Next, we prove the existence and uniqueness of y' in (3.9). Let $U_\varepsilon \coloneqq \{\eta \in \mathbb{C} : |\eta| < 2\varepsilon\}$ and consider the map

$$
\begin{aligned}
F\colon D_{\check{r}}(\mathsf{y}) \times U_\varepsilon &\longrightarrow \qquad \mathbb{C}^d \\
(y,\eta) \quad &\longmapsto K_y(y) + \eta \widetilde{K}_{y'}(y) - K_y(\mathsf{y}).
\end{aligned}
$$

Then

- $F(\mathsf{y},0) = 0, \quad F_y(\mathsf{y},0)^{-1} = K_{yy}(\mathsf{y})^{-1} = T.$
- For any $(y,\eta) \in D_{\check{r}}(\mathsf{y}) \times U_\varepsilon$,

$$
\begin{aligned}
\|\mathbb{1}_d - TF_y(y,\eta)\| &\leqslant \|\mathbb{1}_d - TK_{yy}\| + |\eta|\ \|T\|\ \|\partial^2_{y'}\widetilde{K}\|_{r/2,\mathsf{y}} \\
&\leqslant d\|T\|\|K_{yyy}\|_{\check{r},\mathsf{y}}\check{r} + 2\varepsilon\,\mathsf{T}\frac{4\mathsf{P}}{r^2} \\
&\leqslant d\mathsf{TK}\frac{\check{r}}{r-\check{r}} + 8\mathsf{T}\frac{\varepsilon\,\mathsf{P}}{r^2} \overset{(3.18)}{\leqslant} d\mathsf{TK}\frac{2\check{r}}{r} + \varepsilon\,\frac{8\mathsf{TP}}{r^2} \\
&\leqslant 2d\mathsf{TK}\frac{\bar{r}}{r} + \frac{1}{2}\varepsilon\,\mathsf{L} \\
&\overset{(3.18)+(3.3)}{\leqslant} \frac{5}{12} + \frac{\sigma}{6} \leqslant \frac{5}{12} + \frac{1}{12} = \frac{1}{2}.
\end{aligned}
$$

- Recalling $\sigma \leqslant \frac{1}{2}$, we have

$$
\begin{aligned}
2\|T\| \|F(\mathsf{y},\cdot)\|_{2\varepsilon\,,0} = 2\|T\| \sup_{U_\varepsilon} |\eta \widetilde{K}_{y'}(\mathsf{y})| \leqslant 2\mathsf{T}\frac{4\varepsilon\,\mathsf{P}}{r} &\leqslant \frac{5\cdot 2^{\nu+d}}{8d}\frac{r}{\mathsf{TK}}\sigma^{\nu+d}\varepsilon\,\mathsf{L} \\
&= 3\cdot 2^d\,(2\sigma)^\nu\,\check{r}\,\sigma^d\varepsilon\,\mathsf{L} \leqslant 3\cdot 2^d\,\check{r}\,\sigma^d\varepsilon\,\mathsf{L} \\
&\overset{(3.3)}{\leqslant} 3\,\check{r}\,(2\sigma)^d\,\frac{\sigma}{3} \leqslant \frac{\check{r}}{2}.
\end{aligned}
\tag{3.21}
$$

Therefore, we can apply the inverse function theorem (Lemma 5). Hence, there exists a function $g\colon U_\varepsilon \to D_{\check{r}}(\mathsf{y})$ such that its graph coincides with $F^{-1}(\{0\})$. In particular, $\mathsf{y}' := g(\varepsilon)$ is the unique $y \in D_{\check{r}}(\mathsf{y})$ satisfying $0 = F(y,\varepsilon) = \partial_y K'(y) - \omega$, i.e., the second part of (3.9). Moreover,

$$
|\mathsf{y}'-\mathsf{y}| \leqslant 2\|T\| \|F(\mathsf{y},\cdot)\|_{2\varepsilon\,,0} \leqslant \frac{8\varepsilon\,\mathsf{TP}}{r} \overset{(3.21)}{\leqslant} 3\cdot 2^d\,\check{r}\,\sigma^d\varepsilon\,\mathsf{L} \leqslant \frac{\check{r}}{2}, \tag{3.22}
$$

so that

$$
D_{\frac{\check{r}}{2}}(\mathsf{y}') \subset D_{\check{r}}(\mathsf{y}). \tag{3.23}
$$

Next, we prove that $\partial_y^2 K'(\mathsf{y}')$ is invertible. Indeed, by Taylor' formula, we have

$$
\begin{aligned}
\partial_y^2 K'(\mathsf{y}') &= K_{yy}(\mathsf{y}) + \int_0^1 K_{yyy}(\mathsf{y} + t\varepsilon\widetilde{y})\cdot\varepsilon\widetilde{y}dt + \varepsilon\widetilde{K}_{yy}(\mathsf{y}') \\
&= T^{-1}\left(\mathbb{1}_d + \varepsilon T\left(\int_0^1 K_{yyy}(\mathsf{y} + t\varepsilon\widetilde{y})\cdot\widetilde{y}dt + \widetilde{K}_{yy}(\mathsf{y}')\right)\right) \\
&=: T^{-1}(\mathbb{1}_d + \varepsilon A),
\end{aligned}
$$

and, by Cauchy's estimate,

$$
\begin{aligned}
\varepsilon\,\|A\| &\leqslant \|T\|\left(d\|K_{yyy}\|_{r/2,\mathsf{y}}\varepsilon\,|\mathsf{y}'-\mathsf{y}| + \varepsilon\,\|\partial_{y'}^2\widetilde{K}\|_{r/2,\mathsf{y}}\right) \\
&\leqslant \|T\|\left(\frac{d\|K_{yy}\|_{r,\mathsf{y}}}{r-\frac{r}{2}}\varepsilon\,|\mathsf{y}'-\mathsf{y}| + \varepsilon\,\|\widetilde{K}_{yy}\|_{r/2,\mathsf{y}}\right) \\
&\overset{(3.22)}{\leqslant} \mathsf{T}\left(\frac{2d\mathsf{K}}{r}\frac{8\varepsilon\,\mathsf{TP}}{r} + \frac{4\varepsilon\,\mathsf{P}}{r^2}\right) \leqslant \frac{4\varepsilon\,\mathsf{TP}}{r^2}(4d\mathsf{TK}+1) \\
&\leqslant \frac{20d\varepsilon\,\mathsf{T}^2\mathsf{KP}}{r^2} \leqslant \frac{1}{2}\varepsilon\,\mathsf{L} \overset{(3.3)}{\leqslant} \frac{\sigma}{6} \leqslant \frac{1}{2}.
\end{aligned}
$$

Hence, $\partial_{y'}^2 K'(\mathsf{y}')$ is invertible with

$$
\partial_{y'}^2 K'(\mathsf{y}')^{-1} = (\mathbb{1}_d + \varepsilon A)^{-1}T = T + \sum_{k\geqslant q1}(-\varepsilon)^k A^k T =: T + \varepsilon\widetilde{T},
$$

and

$$
\varepsilon\,\|\widetilde{T}\| \leqslant \varepsilon\,\frac{\|A\|}{1-\varepsilon\,\|A\|}\|T\| \leqslant 2\varepsilon\,\|A\|\|T\| \leqslant \varepsilon\,\mathsf{LT} \leqslant 2\frac{\sigma}{6}\mathsf{T} = \mathsf{T}\frac{\sigma}{3}.
$$

Next, we prove the estimate on P_+. We have

$$
\varepsilon\,\|g_x\|_{\bar{r},\bar{s},\mathsf{y}} \leqslant \varepsilon\,\mathsf{C}_1\frac{\mathsf{P}}{\alpha}\sigma^{-(\tau+d+1)} \leqslant \varepsilon\,\frac{r}{3}\mathsf{L} \overset{(3.3)}{\leqslant} \frac{r}{3}\frac{\sigma}{3} \leqslant \frac{r}{3},
$$

so that, for any $(y',x) \in D_{\bar{r},\bar{s}}(\mathsf{y})$,

$$
|y' + \varepsilon g_x(y',x) - \mathsf{y}| \leqslant \bar{r} + \frac{r}{3} < \frac{r}{8d} + \frac{r}{3} < \frac{2r}{3} < r\,,
$$

and thus

$$\|P^{(1)}\|_{\bar r,\bar s,\mathsf{y}} \leqslant d^2\|K_{yy}\|_{r,\mathsf{y}}\|g_x\|^2_{\bar r,\bar s,\mathsf{y}} \leqslant d^2\mathsf{K}\left(\mathsf{C}_1\frac{\mathsf{P}}{\alpha}\sigma^{-(\nu+d)}\right)^2 = d^2\mathsf{C}_1^2\frac{\mathsf{K}\mathsf{P}^2}{\alpha^2}\sigma^{-2(\nu+d)},$$

$$\|P^{(2)}\|_{\bar r,\bar s,\mathsf{y}} \leqslant d\|P_y\|_{\frac{5r}{6},\bar s,\mathsf{y}}\|g_x\|_{\bar r,\bar s,\mathsf{y}} \leqslant d\frac{6\mathsf{P}}{r}\mathsf{C}_1\frac{\mathsf{P}}{\alpha}\sigma^{-(\nu+d)} = 6d\mathsf{C}_1\frac{\mathsf{P}^2}{\alpha r}\sigma^{-(\nu+d)},$$

and by Fourier estimates (Lemma 4**(ii)**), we have

$$\begin{aligned}
\varepsilon\,\|P^{(3)}\|_{\bar r,s-\frac{\sigma}{2},\mathsf{y}} &\leqslant \sum_{|n|_1>\kappa}\|P_n\|_{\bar r,\mathsf{y}}\,e^{(s-\frac{\sigma}{2})|n|_1} \leqslant \mathsf{P}\sum_{|n|_1>\kappa}e^{-\frac{\sigma|n|_1}{2}}\\
&\leqslant \mathsf{P}\,e^{-\frac{\kappa\sigma}{4}}\sum_{|n|_1>\kappa}e^{-\frac{\sigma|n|_1}{4}} \leqslant \mathsf{P}\,e^{-\frac{\kappa\sigma}{4}}\sum_{|n|_1>0}e^{-\frac{\sigma|n|_1}{4}}\\
&= \mathsf{P}\,e^{-\frac{\kappa\sigma}{4}}\left(\left(\sum_{k\in\mathbb{Z}}e^{-\frac{\sigma|k|}{4}}\right)^d - 1\right) = \mathsf{P}\,e^{-\frac{\kappa\sigma}{4}}\left(\left(1+\frac{2\,e^{-\frac{\sigma}{4}}}{1-\,e^{-\frac{\sigma}{4}}}\right)^d - 1\right)\\
&= \mathsf{P}\,e^{-\frac{\kappa\sigma}{4}}\left(\left(1+\frac{2}{e^{\frac{\sigma}{4}}-1}\right)^d - 1\right) \leqslant \mathsf{P}\,e^{-\frac{\kappa\sigma}{4}}\left(\left(1+\frac{2}{\frac{\sigma}{4}}\right)^d - 1\right)\\
&\leqslant \sigma^{-d}\mathsf{P}\,e^{-\frac{\kappa\sigma}{4}}\left((\sigma+8)^d-\sigma^d\right) \leqslant d8^d\sigma^{-d}\mathsf{P}\,e^{-\frac{\kappa\sigma}{4}}\\
&= \mathsf{C}_2\sigma^{-d}\mathsf{P}\,e^{-\lambda} \overset{(3.2)}{\leqslant} \mathsf{C}_2\sigma^{-d}\mathsf{P}\sigma^{-(2\nu+d)}\frac{\varepsilon\,\mathsf{P}\mathsf{K}}{\alpha^2} = \mathsf{C}_2\mathsf{P}\frac{\varepsilon\,\mathsf{P}\mathsf{K}}{\alpha^2}\sigma^{-2(\nu+d)}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\|P_+\|_{\bar r,\bar s,\mathsf{y}} &\leqslant \|P^{(1)}\|_{\bar r,\bar s,\mathsf{y}} + \|P^{(2)}\|_{\bar r,\bar s,\mathsf{y}} + \|P^{(3)}\|_{\bar r,\bar s,\mathsf{y}}\\
&\leqslant d^2\mathsf{C}_1^2\frac{\mathsf{K}\mathsf{P}^2}{\alpha^2}\sigma^{-2(\nu+d)} + 6d\mathsf{C}_1\frac{\mathsf{P}^2}{\alpha r}\sigma^{-(\nu+d)} + \mathsf{C}_2\mathsf{P}\frac{\varepsilon\,\mathsf{P}\mathsf{K}}{\alpha^2}\sigma^{-2(\nu+d)}\\
&= \left(d^2\mathsf{C}_1^2 r\mathsf{K} + 6d\mathsf{C}_1\alpha\sigma^{\nu+d} + \mathsf{C}_2 r\mathsf{K}\right)\frac{\mathsf{P}^2}{\alpha^2 r}\sigma^{-2(\tau+d+1)}\\
&\leqslant \left(d^2\mathsf{C}_1^2 + 6d\mathsf{C}_1 + \mathsf{C}_2\right)\max\{\alpha, r\mathsf{K}\}\,\frac{\mathsf{P}^2}{\alpha^2 r}\sigma^{-2(\tau+d+1)}\\
&\leqslant \frac{\mathsf{C}_3}{\sqrt{2}}\max\left\{1,\frac{\alpha}{r\mathsf{K}}\right\}\frac{\mathsf{P}^2\mathsf{K}}{\alpha^2}\sigma^{-2(\nu+d)} \leqslant \mathsf{L}\mathsf{P}.
\end{aligned}$$

Finally, we prove that, given $y'\in D_{\bar r}(\mathsf{y})$, the function $\psi_\varepsilon(x) = x+\varepsilon g_{y'}(y',x)$ has an analytic inverse[11]. Consider the Banach space

$$\mathcal{B} := \left\{u\in C^1(\mathbb{T}^d_{s'},\mathbb{C}^d) : \|u\|_{s',1} := \max\left\{\|u\|_{s'}, \|\partial_x u\|_{s'}\right\} \leqslant \overline{\mathsf{L}}\right\}.$$

For any $u\in\mathcal{B}$ and any $x'\in\mathbb{T}^d_{s'}$, we have $\operatorname{Im}\left(x'+\varepsilon u(x')\right) \leqslant s'+\varepsilon\,\|u\|_{s'} \leqslant s'+\varepsilon\,\overline{\mathsf{L}} \overset{(3.3)}{\leqslant} s'+\sigma/6 = s''$. Hence, the functional $f\colon \mathcal{B}\ni u\mapsto -g_{y'}(y',\mathrm{id}+\varepsilon u)$ is well-defined and smooth. Moreover, for any $u\in\mathcal{B}$,

$$\|f(u)\|_{s'} \leqslant \|g_{y'}\|_{W''} \leqslant \overline{\mathsf{L}},\quad \|\partial_x(f(u))\|_{s'} \leqslant \|g_{y'x}\|_{W''}\cdot|\varepsilon|\|\partial_x u\|_{s'} \leqslant \overline{\mathsf{L}}\cdot|\varepsilon|\overline{\mathsf{L}} \overset{(3.3)}{\leqslant} \overline{\mathsf{L}}\cdot\frac{\sigma}{6} < \overline{\mathsf{L}}.$$

[11] Observe that $\psi_\varepsilon(id+\varepsilon u) = id$ is equivalent to $u = -g_{y'}(y', id+\varepsilon u)$, i. e., u is a fixed point of the map $u\mapsto -g_{y'}(y', id+\varepsilon u)$.

Thus, $f\colon \mathcal{B}\to\mathcal{B}$. Furthermore, for any $u_1,u_2\in\mathcal{B}$,

$$\|f(u_1)-f(u_2)\|_{s',1}\leqslant(1+d^2\varepsilon\,\overline{\mathsf{L}})\varepsilon\,\overline{\mathsf{L}}\cdot\|u_1-u_2\|_{s',1}\overset{(3.3)}{\leqslant}2\frac{\sigma}{3d^2}\cdot\|u_1-u_2\|_{s',1}<\frac{1}{2}\|u_1-u_2\|_{s',1},$$

Hence, f is a contraction. Therefore, by the Banach – Caccioppoli fixed-point theorem, f has a unique fixed point $\widetilde{\varphi}_\varepsilon\in\mathcal{B}$; $\widetilde{\varphi}_\varepsilon$ is obtained as the uniform limit $\lim\limits_n f^n(0)$ (as $0\in\mathcal{B}$). Thus, as $f^0=f$ is real-analytic on $D_{\bar r}(\mathsf{y})\times\mathbb{T}^d_{s'}$, by Weierstrass's theorem on the uniform convergence of analytic functions, $\widetilde{\varphi}_\varepsilon$ is real-analytic on $D_{\bar r,s'}(\mathsf{y})$. The rest of the claims on ϕ' and P' are then obvious. □

3.2. Arnold's Scheme: Iteration

Let d, τ, H, K, P, T, ε, α, r, s, s_*, P, K, T, θ, ϵ be as in Theorem A. Set $K_0:=K$, $P_0:=P$, $H_0:=H$. Then, starting from H_0, we shall iterate infinitely many times Lemma 1. The very first step being quite different from all the others, it shall be done separately.

Before starting, let us give some definitions[12].

$$\begin{aligned}
&\epsilon_0:=\epsilon\,,\quad \theta_0:=\theta\,,\quad r_0:=r\,,\quad \mathsf{T}_0:=\mathsf{T}\,,\quad \mathsf{K}_0:=\mathsf{K}\,,\quad \mathsf{P}_0:=\mathsf{P}\,,\\
&\sigma_0:=(s-s_*)/2\,,\quad \lambda_0:=\log\epsilon^{-1}\,,\quad \kappa_0:=4\sigma_0^{-1}\lambda_0\,,\\
&\mathsf{C}_5:=\frac{3\cdot 2^5 d}{5}\,,\quad \mathsf{C}_6:=\max\left\{2^{2\nu}\,,\,\mathsf{C}_5\right\},\quad \mathsf{C}_7:=3d\cdot 2^{6\nu+2d+3}\sqrt{2}\max\left\{640d^2,\,\mathsf{C}_4\right\}\,,\\
&\mathsf{C}_8:=\left(2^{-d}\mathsf{C}_6\right)^{\frac18}\,,\quad \mathsf{C}_9:=3\max\left\{80d\sqrt{2}\,,\,\mathsf{C}_4\right\},\\
&\lambda_*:=\mathsf{C}_7\,\sigma_0^{-(4\nu+2d+1)}\lambda_0^{2\nu}\,\theta^2,\quad \theta_*:=2^{2\nu+2d+1}\,\mathsf{C}_6^2\,\theta^2.\\
&\hat\epsilon_0:=\mathsf{C}_9\,\sigma_0^{-2(\nu+d)-1}\epsilon_0\,\theta_0,\qquad \mathsf{P}_1:=\frac{\hat\epsilon_0\mathsf{P}_0}{\varepsilon}.
\end{aligned}$$

We also set, for $j\geqslant 0$,

$$\begin{aligned}
&\sigma_j:=\frac{\sigma_0}{2^j}\,,\quad s_{j+1}:=s_j-\sigma_j=s_*+\frac{\sigma_0}{2^j}\,,\quad \bar s_j:=s_j-\frac{2\sigma_i}{3}\,,\quad \kappa_j:=4^j\kappa_0,\\
&\mathsf{K}_{j+1}:=\mathsf{K}_0\prod_{k=0}^{j}(1+\frac{\sigma_k}{3})\leqslant\mathsf{K}_0\,e^{\frac{2\sigma_0}{3}}\leqslant\mathsf{K}_0\sqrt{2}\,,\ \mathsf{T}_{j+1}:=\mathsf{T}_0\prod_{k=0}^{j}(1+\frac{\sigma_k}{3})\leqslant\mathsf{T}_0\sqrt{2},\\
&r_{j+1}:=\frac12\min\left\{\frac{\alpha}{2d\sqrt{2}\mathsf{K}_0\kappa_j^\nu},\,\frac{5}{48d}\frac{r_j}{\theta_0}\right\}\,,\quad \mathsf{W}_j:=\operatorname{diag}\left(\max\left\{\frac{\mathsf{K_j}}{\alpha}\,,\frac{1}{r_j}\right\}\mathbb{1}_d,\mathbb{1}_d\right),\\
&\mathsf{L}_j:=\mathsf{P}_i\max\left\{\frac{80d\sqrt{2}\;\mathsf{T}_0\;\theta_0}{r_j^2}\sigma_j^{-(\nu+d)}\,,\,\mathsf{C}_4\max\left\{1,\frac{\alpha}{r_j\mathsf{K}_j}\right\}\frac{\mathsf{K}_0}{\alpha^2}\sigma_j^{-2(\nu+d)}\right\}.
\end{aligned}$$

Observe that

$$\mathsf{W}_0=\operatorname{diag}\left(\mathsf{K}\alpha^{-1}\mathbb{1}_d,\mathbb{1}_d\right)\,,\quad s_j\downarrow s_*\,,\quad r_j\downarrow 0\,,\quad e\,\epsilon_0\leqslant\hat\epsilon_0\,.$$

Note also that, since $\hat\epsilon_0$ is proportional to ε, P_1 *is independent of* ε.

3.2.1. First step

Lemma 2. *Assume*

$$\alpha\leqslant\frac{r_0}{\mathsf{T}_0}\qquad\textit{and}\qquad\hat\epsilon_0\leqslant 1\,.\tag{3.24}$$

[12] Recall the definitions of ν and C_4 given at the beginning of § 3.1.

Then there exist $y_1 \in D_{r_0}(y_0)$ *and a real-analytic symplectic transformation*

$$\phi_0 : D_{r_1,s_1}(y_1) \to D_{r_0,s_0}(y_0)\ , \tag{3.25}$$

such that, for $H_1 \coloneqq H_0 \circ \phi_0$ *, we have*

$$\begin{cases} H_1 =: K_1 + \varepsilon^2 P_1\ , \\ \partial_{y_1} K_1(y_1) = \omega\ , \quad \det \partial^2_{y_1} K_1(y_1) \neq 0 \end{cases} \tag{3.26}$$

and

$$|y_1 - y_0| \leqslant \frac{8\varepsilon\, \mathsf{T}_0 \mathsf{P}_0}{r_0}\ , \tag{3.27}$$

$$\|K_1\|_{r_1/4,y_1} \leqslant \mathsf{K}_1\ , \qquad \|T_1\| \leqslant \mathsf{T}_1\ , \qquad T_1 \coloneqq \partial^2_{y_1} K_1(y_1)^{-1}\ , \tag{3.28}$$

$$\varepsilon^2 \|P_1\|_{r_1,s_1,y_1} \leqslant \varepsilon^2 \mathsf{P}_1\ , \tag{3.29}$$

$$\max\left\{ \|\mathsf{W}_0(\phi_0 - \mathrm{id})\|_{r_1,s_1,y_1}\ ,\ \|\partial_x \pi_2(\phi_0 - \mathrm{id})\|_{s_1} \right\} \leqslant d^{-2} \sigma_0^{d-1}\, \varepsilon\, \mathsf{L}_0\ . \tag{3.30}$$

Proof. Since

$$\kappa_0 \overset{(3.24)}{\geqslant} 4\sigma_0^{-1} \geqslant 8 \tag{3.31}$$

and

$$\frac{\alpha}{2d\sqrt{2}\mathsf{K}_0 k_0^\nu} \overset{(3.24)+(3.31)}{\leqslant} \frac{1}{2d\cdot 8^\nu \sqrt{2}\mathsf{K}_0} \frac{r_0}{\mathsf{T}_0} < \frac{5}{48d}\frac{r_0}{\theta_0}\ ,$$

we get

$$r_1 = \frac{1}{2}\min\left\{ \frac{\alpha}{2d\sqrt{2}\mathsf{K}_0 \kappa_0^\nu}\ ,\ \frac{5}{48d}\frac{r_0}{\theta_0} \right\} = \frac{\alpha}{4d\sqrt{2}\mathsf{K}_0 \kappa_0^\nu}\ . \tag{3.32}$$

Thus,

$$\begin{aligned} \varepsilon\, \mathsf{L}_0(3\sigma_0^{-1}) &\leqslant 3\varepsilon\, \mathsf{P}_0 \max\left\{ \frac{80d\sqrt{2}\ \mathsf{T}_0\ \theta_0}{r_0^2} \sigma_0^{-(\nu+d)}\ ,\ \mathsf{C}_4 \max\left\{1, \frac{\alpha}{r_0 \mathsf{K}_0}\right\} \frac{\mathsf{K}_0}{\alpha^2} \sigma_0^{-2(\nu+d)} \right\} \sigma_0^{-1} \\ &\leqslant 3\max\left\{ 80d\sqrt{2}\ \theta_0 \frac{\alpha\ \mathsf{T}_0}{r_0}\frac{\alpha}{r_0\mathsf{K}_0}\ ,\ \mathsf{C}_4 \max\left\{1, \frac{\alpha}{r_0\mathsf{K}_0}\right\}\right\} \sigma_0^{-2(\nu+d)-1} \frac{\mathsf{K}_0 \varepsilon\, \mathsf{P}_0}{\alpha^2} \\ &\overset{(3.24)}{\leqslant} 3\max\left\{80d\sqrt{2},\ \mathsf{C}_4\right\} \sigma_0^{-2(\nu+d)-1} \epsilon_0\ \theta_0 = \hat{\epsilon}_0 \overset{(3.24)}{\leqslant} 1. \end{aligned} \tag{3.33}$$

Therefore, Lemma 1 implies Lemma 2. □

3.2.2. Subsequent steps, iteration and convergence

For $j \geqslant 1$, define

$$\epsilon_j \coloneqq \frac{\mathsf{K}_0\, \varepsilon^{2^j} \mathsf{P}_j}{\alpha^2}, \qquad \mathsf{P}_{j+1} \coloneqq \lambda_* \theta_*^{j-1} \frac{\mathsf{K}_0 {\mathsf{P}_j}^2}{\alpha^2}, \qquad \hat{\epsilon}_j \coloneqq \lambda_*\ \theta_*^j\ \epsilon_j.$$

Thus, for any $j \geqslant 1$, one has

$$\begin{aligned} \hat{\epsilon}_{j+1} = \lambda_*\ \theta_*^{j+1}\ \epsilon_{j+1} &= \lambda_*\ \theta_*^{j+1} \frac{\mathsf{K}_0 \varepsilon^{2^{j+1}} \mathsf{P}_{j+1}}{\alpha^2} = \lambda_*\ \theta_*^{j+1} \frac{\mathsf{K}_0 \varepsilon^{2^{j+1}}}{\alpha^2}\ \lambda_* \theta_*^{j-1} \frac{\mathsf{K}_0 {\mathsf{P}_j}^2}{\alpha^2} \\ &= \left(\lambda_*\ \theta_*^j\ \epsilon_j\right)^2 = \hat{\epsilon}_j^2, \end{aligned}$$

i. e.,

$$\hat{\epsilon}_j = \hat{\epsilon}_1^{2^{j-1}}.$$

Once the first step is completed, all the following steps do not need any other condition. Actually, the first condition in (3.24) is no longer necessary and the second condition needs to be strengthened merely a little bit more. To be precise, the following holds.

Lemma 3. *Assume* (3.26) ÷ (3.29) *and*

$$\mathsf{C}_8\ \theta_0^{\frac{1}{8}}\ \hat{\epsilon}_1 < 1\ . \tag{3.34}$$

Then one can construct a sequence of symplectic transformations

$$\phi_{j-1} : D_{r_j,s_j}(y_j) \to D_{r_{j-1},s_{j-1}}(y_{j-1}), \qquad j \geqslant 2 \tag{3.35}$$

so that

$$H_j := H_{j-1} \circ \phi_{j-1} =: K_j + \varepsilon^{2^j} P_j \tag{3.36}$$

converges uniformly.
More precisely, $\varepsilon^{2^{j-1}} P_{j-1}$, $\phi^{j-1} := \phi_1 \circ \phi_2 \circ \cdots \circ \phi_{j-1}$, K_{j-1}, y_{j-1} *converge uniformly on* $\{y_*\} \times \mathbb{T}^d_{s_*}$ *to, respectively,* 0, ϕ^*, K_*, y_* *and* $H_1 \circ \phi^* = K_*$ *with* ϕ^* *real-analytic for* $x \in \mathbb{T}^d_{s_*}$ *and* $\det \partial_y^2 K_*(y_*) \neq 0$.
Finally, the following estimates hold for any $i \geqslant 1$*:*

$$\varepsilon^{2^i} \|P_i\|_{r_i,s_i,y_i} \leqslant \varepsilon^{2^i} \mathsf{P}_i\ , \tag{3.37}$$

$$|y_{i+1} - y_i| \leqslant \frac{8\sqrt{2}\mathsf{T}_0 \varepsilon^{2^i} \mathsf{P}_i}{r_i}\ , \tag{3.38}$$

$$|\mathsf{W}_1(\phi^* - \mathrm{id})| \leqslant \frac{2\sigma_0^d\ \hat{\epsilon}_1}{3d^2\ \theta_*} \qquad \textit{on} \quad \{y_*\} \times \mathbb{T}^d_{s_*}\ . \tag{3.39}$$

Remark 1. Notice that P_1 is actually *independent of* ε (and, in particular, of $\log \epsilon^{-1}$), while P_j for $j \geqslant 2$ does depend on $\log \epsilon^{-1}$ through λ_*. This is a crucial point, which allows us, in the end, to get optimal bounds on the displacement of the persistent invariant torus from the unperturbed one.

Proof. First of all, notice that, for any $i \geqslant 1$,

$$\begin{aligned} r_{i+1} &= \min\left\{\frac{\alpha}{4d\sqrt{2}\mathsf{K}_0\kappa_i^\nu}\, ,\, \frac{5}{96d}\frac{r_i}{\theta_0}\right\} = \min\left\{\frac{r_1}{4^{i\nu}}\, ,\, \frac{5}{96d\theta_0} r_i\right\} \\ &= \min\left\{\frac{r_1}{4^{\nu i}}\, ,\, \frac{5}{96d\theta_0}\frac{r_1}{4^{\nu(i-1)}}\, ,\, \left(\frac{5}{96d\theta_0}\right)^2 r_{i-1}\right\} \\ &\ \ \vdots \\ &= \min\left\{\frac{r_1}{4^{\nu i}}\, ,\, \frac{5}{96d\theta_0}\frac{r_1}{4^{\nu(i-1)}}\, ,\cdots ,\, \left(\frac{5}{96d\theta_0}\right)^i r_1\right\} \\ &= \frac{r_1}{4^{\nu i}} \min\left\{\left(\frac{5\cdot 4^\nu}{96d\theta_0}\right)^0 ,\cdots ,\, \left(\frac{5\cdot 4^\nu}{96d\theta_0}\right)^i\right\} \\ &= \frac{r_1}{4^{\nu i}} \min{}^i \left\{\frac{5\cdot 4^\nu}{96d\theta_0}\, ,\, 1\right\} = r_1 \min{}^i\left\{\frac{1}{2^{2\nu}}\, ,\, \frac{5}{96d\theta_0}\right\} = \frac{r_1}{\mathbf{a}_1^i}\, , \end{aligned}$$

where

$$\mathbf{a}_1 := \max\left\{2^{2\nu}\, ,\, \frac{96d\theta_0}{5}\right\} \leqslant \max\left\{2^{2\nu}\, ,\, \frac{96d}{5}\right\} \cdot \theta_0 = \mathsf{C}_6\ \theta_0\ . \tag{3.40}$$

For a given $j \geqslant 2$, let $(\mathscr{P}^j)$ be the following assertion:

there exist $j-1$ *symplectic transformations*[13)]

$$\phi_i : D_{r_{i+1},s_{i+1}}(y_{i+1}) \to D_{2r_i/3,\bar{s}_i}(y_i), \quad \textit{for} \quad 1 \leqslant i \leqslant j-1, \tag{3.41}$$

[13)] Compare (3.4).

and $j-1$ *Hamiltonians* $H_{i+1} = H_i \circ \phi_i = K_{i+1} + \varepsilon^{2^{i+1}} P_{i+1}$ *real–analytic on* $D_{r_{i+1},s_{i+1}}(y_{i+1})$ *such that, for any* $1 \leqslant i \leqslant j-1$,

$$\begin{cases} \|\partial_y^2 K_i\|_{r_i,y_i} \leqslant \mathsf{K}_i \,, \quad \|T_i\| \leqslant \mathsf{T}_i \,, \quad \partial_y K_i(y_i) = \omega \,, \quad \partial_y^2 K_i(y_i) \neq 0 \,, \\ \|P_i\|_{r_i,s_i,y_i} \leqslant \mathsf{P}_i \,, \quad \kappa_i \geqslant 4\sigma_i^{-1} \log\left(\sigma_i^{2\nu+d} \epsilon_i^{-1}\right) \,, \quad \varepsilon^{2^i} \mathsf{L}_i \leqslant \dfrac{\sigma_i}{3} \end{cases} \tag{3.42}$$

and

$$\begin{cases} \partial_y K_{i+1}(y_{i+1}) = \omega \,, \quad \partial_y^2 K_{i+1}(y_{i+1}) \neq 0 \,, \quad |y_{i+1} - y_i| \leqslant \dfrac{8\sqrt{2}\mathsf{T}_0 \varepsilon^{2^i} \mathsf{P}_i}{r_i} \,, \\ \|T_{i+1}\| \leqslant \|T_i\| + \mathsf{T}_i \varepsilon^{2^i} \mathsf{L}_i \,, \quad \|K_{i+1}\|_{r_{i+1},y_{i+1}} \leqslant \|K_i\|_{r_i,y_i} + \varepsilon^{2^i} \mathsf{P}_i \,, \\ \|\partial_y^2 K_{i+1}\|_{r_{i+1},y_{i+1}} \leqslant \|\partial_y^2 K_i\|_{r_i,y_i} + \mathsf{K}_i \varepsilon^{2^i} \mathsf{L}_i \,, \\ \max\left\{ \|\mathsf{W}_i(\phi_i - \mathrm{id})\|_{r_{i+1},s_{i+1},y_{i+1}} \,, \|\partial_x \pi_2(\phi_i - \mathrm{id})\|_{s_{i+1}} \right\} \leqslant d^{-2}\, \sigma_i^{d-1}\, \varepsilon^{2^i} \mathsf{L}_i \,, \\ \|P_{i+1}\|_{r_{i+1},s_{i+1},y_{i+1}} \leqslant \mathsf{P}_i \mathsf{L}_i \,. \end{cases} \tag{3.43}$$

Assume $(\mathscr{P}^j)$ for some $j \geqslant 2$ and let us check $(\mathscr{P}^{j+1})$. Fix $1 \leqslant i \leqslant j-1$. Then

$$\|\partial_y^2 K_{i+1}\|_{r_{i+1},y_{i+1}} \overset{(3.43)}{\leqslant} \|\partial_y^2 K_i\|_{r_i,y_i} + \mathsf{K}_i \varepsilon^{2^i} \mathsf{L}_i \overset{(3.42)}{\leqslant} \mathsf{K}_i + \mathsf{K}_i \frac{\sigma_i}{3} = \mathsf{K}_{i+1} < \mathsf{K}_0 \sqrt{2}$$

and, similarly,

$$\|T_{i+1}\| \leqslant \mathsf{T}_{i+1},$$

which proves the two first relations in (3.42) for $i = j$. Also,

$$\frac{\alpha}{r_i \mathsf{K}_i} > \frac{\alpha}{r_1 \mathsf{K}_0 \sqrt{2}} = 4d\kappa_0^\nu \overset{(3.31)}{>} 1 \,, \tag{3.44}$$

so that

$$\begin{aligned}
\varepsilon^{2^i} \mathsf{L}_i (3\sigma_i^{-1}) &= 3\varepsilon^{2^i} \mathsf{P}_i \max\left\{ \frac{80d\sqrt{2}\mathsf{T}_0\theta_0}{r_i^2} \sigma_i^{-(\nu+d)} \,, \mathsf{C}_4 \max\left\{1, \frac{\alpha}{r_i \mathsf{K}_i}\right\} \frac{\mathsf{K}_0}{\alpha^2} \sigma_i^{-2(\nu+d)} \right\} \sigma_i^{-1} \\
&\overset{(3.44)}{\leqslant} 3\varepsilon^{2^i} \mathsf{P}_i \max\left\{ \frac{80d\sqrt{2}\mathsf{T}_0\theta_0}{r_i^2} \,, \mathsf{C}_4 \frac{1}{\alpha r_i} \right\} \sigma_i^{-2(\nu+d)-1} \\
&= 3\max\left\{ 80d\sqrt{2}\mathsf{T}_0\theta_0 \frac{\alpha}{r_i} \,, \mathsf{C}_4 \right\} \sigma_i^{-2(\nu+d)-1} \frac{\varepsilon^{2^i} \mathsf{P}_i}{\alpha r_i} \\
&= 3\max\left\{ 640 d^2 \theta_0^2 \mathbf{a}^{i-1} \kappa_0^\nu \,, \mathsf{C}_4 \right\} \sigma_i^{-2(\nu+d)-1} \frac{\varepsilon^{2^i} \mathsf{P}_i}{\alpha^2} 4d\sqrt{2}\mathsf{K}_0 \kappa_0^\nu \mathbf{a}^{i-1} \\
&\overset{(3.31)}{\leqslant} 12d\sqrt{2} \max\left\{ 640d^2 \,, \mathsf{C}_4 \right\} \sigma_i^{-2(\nu+d)-1} \frac{\mathsf{K}_0 \varepsilon^{2^i} \mathsf{P}_i}{\alpha^2} \theta_0^2 \mathbf{a}^{2(i-1)} \kappa_0^{2\nu} \\
&\overset{(3.40)}{\leqslant} 12d\sqrt{2} \max\left\{ 640d^2 \,, \mathsf{C}_4 \right\} \sigma_i^{-2(\nu+d)-1} \frac{\mathsf{K}_0 \varepsilon^{2^i} \mathsf{P}_i}{\alpha^2} \theta_0^{2i} \mathsf{C}_6^{2(i-1)} \kappa_0^\nu \\
&= 3d \cdot 2^{6\nu+2d+3} \sqrt{2} \max\left\{ 640d^2 \,, \mathsf{C}_4 \right\} \sigma_0^{-(4\nu+2d+1)} \left(2^{2\nu+2d+1} \mathsf{C}_6^2 \theta_0^2 \right)^{i-1} \\
&\quad \times \frac{\mathsf{K}_0 \varepsilon^{2^i} \mathsf{P}_i}{\alpha^2} \left(\log \epsilon_0^{-1}\right)^{2\nu} \theta_0^2 \\
&\leqslant \mathsf{C}_7 \sigma_0^{-(4\nu+2d+1)} \left(\log \epsilon_0^{-1}\right)^{2\nu} \theta_0^2\, \theta_*^{i-1} \frac{\mathsf{K}_0 \varepsilon^{2^i} \mathsf{P}_i}{\alpha^2} = \lambda_*\, \theta_*^{i-1} \epsilon_i = \frac{\hat{\epsilon}_i}{\theta_*} = \frac{\hat{\epsilon}_1^{2^{i-1}}}{\theta_*} \\
&\overset{(3.34)}{\leqslant} \frac{1}{\theta_*} < 1 \,.
\end{aligned}$$

Moreover,

$$\varepsilon^{2^i}\mathsf{L}_i < \lambda_*\,\theta_*^{i-1}\epsilon_i\ .$$

Thus, by the last relation in (3.43), for any $1\leqslant i\leqslant j-1$,

$$\varepsilon^{2^{i+1}}\|P_{i+1}\|_{r_{i+1},s_{i+1},y_{i+1}}\leqslant \varepsilon^{2^i}\mathsf{L}_i\,\varepsilon^{2^i}\mathsf{P}_i<\lambda_*\theta_*^{i-1}\,\epsilon_i\,\varepsilon^{2^i}\mathsf{P}_i=\varepsilon^{2^{i+1}}\mathsf{P}_{i+1}\ ,$$

which proves the fourth relation in (3.42) for $i=j$. Furthermore, by exactly the same computation as above, one gets

$$\varepsilon^{2^{i+1}}\mathsf{L}_{i+1}(3\sigma_{i+1}^{-1})\leqslant\frac{\hat\epsilon_{i+1}}{\theta_*}=\frac{\hat\epsilon_1^{2^i}}{\theta_*}<1\ ,$$

which proves the last relation in (3.42) for $i=j$. It remains only to check that the fifth relation in (3.42) holds as well for $i=j$ in order to apply Lemma 1 to H_i, $1\leqslant i\leqslant j$ and get (3.43) and, consequently, $(\mathscr{P}^{j+1})$. In fact, we have[14)]

$$\lambda_*\,\theta_*\,\epsilon_0^2<\lambda_*\,\theta_*\,\epsilon_0\,\hat\epsilon_0=\hat\epsilon_1\leqslant\mathsf{C}_7\,\sigma_0^{-(4\nu+2d+1)}\,\theta_*\,\theta_0^2\,\hat\epsilon_0\ ,\tag{3.45}$$

so that

$$\begin{aligned}4\sigma_j^{-1}\log\left(\sigma_j^{2\nu+d}\epsilon_j^{-1}\right)&\leqslant 4\sigma_j^{-1}\log\left(\epsilon_j^{-1}\right)=4\sigma_j^{-1}\log\left(\lambda_*\theta_*^j\hat\epsilon_1^{-2^{j-1}}\right)\\&\overset{(3.45)}{\leqslant}4\sigma_j^{-1}\log\left(\lambda_*\theta_*^j(\lambda_*\theta_*\epsilon_0^2)^{-2^{j-1}}\right)\leqslant 4\sigma_j^{-1}\log\left(\epsilon_0^{-2^j}\right)\\&=4^j\cdot 4\sigma_0^{-1}\log\left(\epsilon_0^{-1}\right)=\kappa_j\ .\end{aligned}$$

To finish the proof of the induction, i.e., to construct an *infinite sequence* of Arnold's transformations satisfying (3.42) and (3.43) *for all* $i\geqslant 1$, one needs only to check $(\mathscr{P}^2)$. Thanks to[15)] $(3.26)\div(3.29)$, we just need to check the two last inequalities in $(3.42)_{i=1}$. But, in fact, this is contained in the above computation. Then we apply Lemma 1 to H_1 to get $(3.41)_{i=1}$ and $(3.43)_{i=1}$, which achieves the proof of $(\mathscr{P}^2)$.

Next, we prove that ϕ^j is convergent by proving that it is a Cauchy sequence. For any $j\geqslant 4$, we have, using again Cauchy's estimate (and noting that $2^{i-1}\geqslant i$, $\forall\, i\geqslant 0$),

$$\begin{aligned}\|\mathsf{W}_{j-1}(\phi^{j-1}-\phi^{j-2})\|_{r_j,s_j,y_j}&=\|\mathsf{W}_{j-1}\phi^{j-2}\circ\phi_{j-1}-\mathsf{W}_{j-1}\phi^{j-2}\|_{r_j,s_j,y_j}\\&\overset{(3.41)}{\leqslant}\|\mathsf{W}_{j-1}D\phi^{j-2}\mathsf{W}_{j-1}^{-1}\|_{2r_{j-1}/3,s_{j-1},y_{j-1}}\,\|\mathsf{W}_{j-1}(\phi_{j-1}-\mathrm{id})\|_{r_j,s_j,y_j}\\&\overset{(3.43)}{\leqslant}\max\left(r_{j-1}\frac{3}{r_{j-1}},\frac{3}{2\sigma_{j-1}}\right)\|\mathsf{W}_{j-1}\phi^{j-2}\|_{r_{j-1},s_{j-1},y_{j-1}}\times\\&\qquad\times\|\mathsf{W}_{j-1}(\phi_{j-1}-\mathrm{id})\|_{r_j,s_j,y_j}\\&=\frac{3}{2\sigma_{j-1}}\|\mathsf{W}_{j-1}\phi^{j-2}\|_{r_{j-1},s_{j-1},y_{j-1}}\,\|\mathsf{W}_{j-1}(\phi_{j-1}-\mathrm{id})\|_{r_j,s_j,y_j}\\&\leqslant\frac12\|\mathsf{W}_{j-1}\phi^{j-2}\|_{r_{j-1},s_{j-1},y_{j-1}}\cdot\sigma_{j-1}^d\left(\varepsilon^{2^{j-1}}\mathsf{L}_{j-1}3\sigma_{i-1}^{-1}\right)\\&\leqslant\frac12\|\mathsf{W}_{j-1}\phi_1\|_{r_2,s_2,y_2}\cdot\sigma_{j-1}^d\,\hat\epsilon_{j-1}\\&\leqslant\frac12\left(\prod_{i=1}^{j-2}\|\mathsf{W}_{i+1}\mathsf{W}_i^{-1}\|\right)\|\mathsf{W}_1\phi_1\|_{r_2,s_2,y_2}\cdot\sigma_{j-1}^d\,\hat\epsilon_{j-1}\end{aligned}$$

[14)] Notice that $(\log t)^{2s}\leqslant t^{1/2}$, $\forall\, t\geqslant e$, $\forall\, s\geqslant 1/4$, so that $\epsilon_0(\log\epsilon_0^{-1})^{2\nu}\overset{(3.34)}{\leqslant}\sqrt{\epsilon_0}\leqslant e^{-1/2}<1$, which in turn proves the *r.h.s.* inequality in (3.45).

[15)] Observe that for $j=2$, $i=1$.

$$
\begin{aligned}
&\overset{(3.44)}{=} \frac{1}{2}\left(\prod_{i=1}^{j-2}\frac{r_i}{r_{i+1}}\right)\|\mathsf{W}_1\phi_1\|_{r_2,s_2,y_2}\cdot\sigma_{j-1}^d\ \hat{\epsilon}_{j-1}\\
&= \frac{r_1}{2r_{j-1}}\|\mathsf{W}_1\phi_1\|_{r_2,s_2,y_2}\cdot\sigma_{j-1}^d\ \hat{\epsilon}_{j-1}\\
&\overset{(3.40)}{\leqslant} \frac{1}{2}\sigma_3^d\ (\mathsf{C}_6\ \theta_0)^2\ \|\mathsf{W}_1\phi_1\|_{r_2,s_2,y_2}\cdot\left(2^{-d}\mathsf{C}_6\ \theta_0\right)^{j-4}\cdot\ \hat{\epsilon}_1^{2^{j-2}}\\
&\leqslant \frac{1}{2}\sigma_3^d\ (\mathsf{C}_6\ \theta_0)^2\ \|\mathsf{W}_1\phi_1\|_{r_2,s_2,y_2}\cdot\left(2^{-d}\mathsf{C}_6\ \theta_0\right)^{2^{j-5}}\cdot\ \hat{\epsilon}_1^{2^{j-2}}\\
&= \frac{1}{2}\sigma_3^d\ (\mathsf{C}_6\ \theta_0)^2\ \|\mathsf{W}_1\phi_1\|_{r_2,s_2,y_2}\cdot\left(\left(2^{-d}\mathsf{C}_6\ \theta_0\right)^{\frac{1}{8}}\hat{\epsilon}_1\right)^{2^{j-2}}\\
&= \frac{1}{2}\sigma_3^d\ (\mathsf{C}_6\ \theta_0)^2\ \|\mathsf{W}_1\phi_1\|_{r_2,s_2,y_2}\cdot\left(\mathsf{C}_8\ \theta_0^{\frac{1}{8}}\ \hat{\epsilon}_1\right)^{2^{j-2}}\ .
\end{aligned}
$$

Therefore, for any $n \geqslant 1$, $j \geqslant q0$,

$$
\begin{aligned}
\|\mathsf{W}_1(\phi^{n+j+1}-\phi^n)\|_{r_{n+j+2},s_{n+j+2},y_{n+j+2}} &\leqslant \sum_{i=n}^{n+j}\|\mathsf{W}_1(\phi^{i+1}-\phi^i)\|_{r_{i+2},s_{i+2},y_{i+2}}\\
&\leqslant \sum_{i=n}^{n+j}\left(\prod_{k=1}^{i}\|\mathsf{W}_k\mathsf{W}_{k+1}^{-1}\|\right)\|\mathsf{W}_{i+1}(\phi^{i+1}-\phi^i)\|_{r_{i+2},s_{i+2},y_{i+2}}\\
&\overset{(3.44)}{=} \sum_{i=n}^{n+j}\prod_{k=1}^{i}\max\left\{1,\frac{r_{k+1}}{r_k}\right\}\|\mathsf{W}_{i+1}(\phi^{i+1}-\phi^i)\|_{r_{i+2},s_{i+2},y_{i+2}}\\
&= \sum_{i=n}^{n+j}\|\mathsf{W}_{i+1}(\phi^{i+1}-\phi^i)\|_{r_{i+2},s_{i+2},y_{i+2}}\\
&\leqslant \frac{1}{2}\sigma_3^d\ (\mathsf{C}_6\ \theta_0)^2\ \|\mathsf{W}_1\phi_1\|_{r_2,s_2,y_2}\cdot\sqrt{\varepsilon}\sum_{i=n}^{n+j}\left(\mathsf{C}_8\ \theta_0^{\frac{1}{8}}\ \hat{\epsilon}_1\right)^{2^i}\ .
\end{aligned}
$$

Hence, by (3.34), ϕ^j converges uniformly on $\{y_*\}\times\mathbb{T}^d_{s_*}$ to some ϕ^*, which is then a real-analytic map in $x\in\mathbb{T}^d_{s_*}$.

To estimate $|\mathsf{W}_0(\phi^*-\mathrm{id})|$ on $\{y_*\}\times\mathbb{T}^d_{s_*}$, observe that, for $i\geqslant 1$,

$$
\sigma_i^d\ \varepsilon^{2^i}\mathsf{L}_i \leqslant \frac{\sigma_0^{d+1}}{3\cdot 2^{i(d+1)}}\ \frac{\hat{\epsilon}_1^{2^{i-1}}}{\theta_*} \leqslant \frac{\sigma_0^{d+1}}{3\cdot 2^{(d+1)i}\theta_*}\hat{\epsilon}_1^i = \frac{(2\sigma_0)^{d+1}}{3\theta_*}\left(\frac{\hat{\epsilon}_1}{2^{d+1}}\right)^{i+1}
$$

and therefore

$$
\sum_{i\geqslant 1}\sigma_i^d\ \varepsilon^{2^i}\mathsf{L}_i \leqslant \frac{(2\sigma_0)^{d+1}}{3\theta_*}\sum_{i\geqslant 1}\left(\frac{\hat{\epsilon}_1}{2^{d+1}}\right)^i \leqslant \frac{2\sigma_0^{d+1}\ \hat{\epsilon}_1}{3\ \theta_*}\ .
$$

Moreover, for any $i\geqslant 1$,

$$
\begin{aligned}
\|\mathsf{W}_1(\phi^i-\mathrm{id})\|_{r_{i+1},s_{i+1},y_{i+1}} &\leqslant \|\mathsf{W}_1(\phi^{i-1}\circ\phi_i-\phi_i)\|_{r_{i+1},s_{i+1},y_{i+1}}+\|\mathsf{W}_1(\phi_i-\mathrm{id})\|_{r_{i+1},s_{i+1},y_{i+1}}\\
&\leqslant \|\mathsf{W}_1(\phi^{i-1}-\mathrm{id})\|_{r_i,s_i,y_i}+\Big(\prod_{j=0}^{i-1}\|\mathsf{W}_j\mathsf{W}_{j+1}^{-1}\|\Big)\|\mathsf{W}_i(\phi_i-\mathrm{id})\|_{r_{i+1},s_{i+1},y_{i+1}}\\
&= \|\mathsf{W}_1(\phi^{i-1}-\mathrm{id})\|_{r_i,s_i,y_i}+\|\mathsf{W}_i(\phi_i-\mathrm{id})\|_{r_{i+1},s_{i+1},y_{i+1}}\\
&= \|\mathsf{W}_1(\phi^{i-1}-\mathrm{id})\|_{r_i,s_i,y_i}+\|\mathsf{W}_i(\phi_i-\mathrm{id})\|_{r_{i+1},s_{i+1},y_{i+1}}\\
&\leqslant \|\mathsf{W}_1(\phi^{i-1}-\mathrm{id})\|_{r_i,s_i,y_i}+d^{-2}\sigma_i^{d-1}\ \varepsilon^{2^i}\mathsf{L}_i\ ,
\end{aligned}
$$

which iterated yields

$$\|\mathsf{W}_1(\phi^i - \mathrm{id})\|_{r_{i+1},s_{i+1},y_{i+1}} \leqslant d^{-2} \sum_{k\geqslant 1} \sigma_k^{d-1}\ \varepsilon^{2^k} \mathsf{L}_k \leqslant \frac{2\sigma_0^d\ \hat{\epsilon}_1}{3d^2\ \theta_*} .$$

Therefore, taking the limit over i completes the proof of (3.39) and hence of Lemma 3. □

CONCLUSION

We can now complete the proof of Theorem A. Let

$$\mathsf{C}_{10} := \Big(2^{-(4\nu+2d+1)} + 2\mathsf{C}_7\Big)\, \mathsf{C}_9/(3d^2) , \quad \mathsf{C}_{11} := \frac{1}{2^{5\nu+3d-2}} + \frac{\mathsf{C}_7\ \mathsf{C}_9}{3\cdot 5\cdot 2^{\nu+2}\cdot d^2\cdot \sqrt{2}} ,$$
$$\mathsf{C}_{12} := 2^{2\nu+2d+1}\, \mathsf{C}_6^2\ \mathsf{C}_7\ \mathsf{C}_8\ \mathsf{C}_9 , \quad \mathsf{C}_{13} := \mathsf{C}_{10} + 2^{-(\nu+1)}\ \mathsf{C}_{11} \quad , \mathsf{C}_{14} := 2^{2(3\nu+2d+1)}\mathsf{C}_{12} ,$$
$$\mathsf{C}_{15} := 18d^3 + 70 , \quad \mathsf{C} := 2^{6\tau+3d+8}\mathsf{C}_{13} , \quad \mathsf{C}_* := \max\Big\{(4\nu\, e^{-1})^{8\nu/3}\mathsf{C}_{14}^{2/3},\ 2\mathsf{C}_{15}\mathsf{C}\Big\} .$$

Observe that

$$(\log t)^{4\nu} \leqslant (4\nu\, e^{-1})^{4\nu}\sqrt{t} , \qquad \forall\, t > 1. \tag{3.46}$$

Then

$$\begin{aligned}\mathsf{C}_8\ \theta_0^{\frac{1}{8}}\ \hat{\epsilon}_1 &= \mathsf{C}_{14}\ \theta^{41/8}\ (s-s_*)^{-2(3\nu+2d+1)}\ \epsilon^2(\log\epsilon^{-1})^{2\nu}\\ &\overset{(3.46)}{\leqslant} (4\nu\, e^{-1})^{4\nu}\ \mathsf{C}_{14}\ \theta^{41/8}\ \epsilon^{3/2}\ (s-s_*)^{-2(3\nu+2d+1)}\\ &< \Big(\mathsf{C}_*\ \theta^4\ (s-s_*)^{-(6\nu+3d+2)}\ \epsilon\Big)^{3/2}\\ &\overset{(2.4)}{\leqslant} 1\end{aligned}$$

and

$$\hat{\epsilon}_0 < \mathsf{C}_*\ \theta^4\ (s-s_*)^{-(6\nu+3d+2)}\ \epsilon \overset{(2.4)}{\leqslant} 1.$$

Hence, (2.4) implies the smallness conditions (3.24) and (3.34). Therefore, Lemmas 2 and 3 hold. Now set $\phi_* := \phi_0\circ\phi^*$ and observe that, uniformly on $\{y_*\}\times\mathbb{T}^d_{s_*}$,

$$\begin{aligned}|\mathsf{W}_0(\phi_*-\mathrm{id})| &\leqslant |\mathsf{W}_0(\phi_0\circ\phi^*-\phi^*)| + |\mathsf{W}_0(\phi^*-\mathrm{id})|\\ &\leqslant \|\mathsf{W}_0(\phi_0-\mathrm{id})\|_{r_1,s_1,y_1} + \|\mathsf{W}_0\mathsf{W}_1^{-1}\|\ |\mathsf{W}_1(\phi^*-\mathrm{id})|\\ &\leqslant \frac{1}{d^2}\sigma_0^d\ \varepsilon\,\mathsf{L}_0 + \frac{2\sigma_0^d}{3d^2\theta_*}\ \hat{\epsilon}_1 \overset{(3.33)+(3.45)}{\leqslant} \frac{\sigma_0^d}{3d^2}\ \hat{\epsilon}_0 + \frac{2\sigma_0^d}{3d^2\theta_*}\mathsf{C}_7\ \sigma_0^{-(4\nu+2d+1)}\ \theta_*\ \theta_0^2\ \hat{\epsilon}_0\\ &\leqslant \left(\frac{1}{3d^2 2^{4\nu+2d+1}} + \frac{2\mathsf{C}_7}{3d^2}\right)\sigma_0^{-(4\nu+d+1)}\ \theta_0^2\ \hat{\epsilon}_0 = \mathsf{C}_{10}\ \sigma_0^{-(6\nu+3d+2)}\ \theta_0^3\ \epsilon_0 =: \gamma .\end{aligned}$$

Moreover, for any $i\geqslant 1$,

$$\begin{aligned}|y_i-y_0| &\leqslant \sum_{j=0}^{i-1}|y_{j-1}-y_j| \overset{(3.27)+(3.38)}{\leqslant} \frac{8\mathsf{T}_0\varepsilon\,\mathsf{P}_0}{r_0} + \sum_{j=1}^{i-1}\frac{8\sqrt{2}\mathsf{T}_0\varepsilon^{2^j}\mathsf{P}_j}{r_j}\\ &\leqslant \frac{8\mathsf{T}_0\varepsilon\,\mathsf{P}_0}{r_0} + \sum_{j\geqslant 1}\frac{r_j}{10d\theta_0}\sigma_j^{\nu+d}\varepsilon^{2^j}\mathsf{L}_j \leqslant \frac{8\mathsf{T}_0\varepsilon\,\mathsf{P}_0}{r_0} + \frac{r_1}{10d\theta_0}\sigma_1^{\nu}\sum_{j\geqslant 1}\sigma_j^d\varepsilon^{2^j}\mathsf{L}_j\\ &\leqslant \frac{8\mathsf{T}_0\varepsilon\,\mathsf{P}_0}{r_0} + \frac{r_1}{10d\theta_0}\sigma_1^{\nu}\ \frac{2\sigma_0^{d+1}\hat{\epsilon}_1}{3\theta_*} \overset{(2.4)+(3.45)}{\leqslant} \mathsf{C}_{11}\ \sigma_0^{-(5\nu+3d+1)}\ \theta_0^2\ \frac{\varepsilon\ \mathsf{P}_0}{\alpha} ,\end{aligned}$$

and then, passing to the limit, we get

$$|y_*-y_0| \leqslant \mathsf{C}_{11}\ \sigma_0^{-(5\nu+3d+1)}\ \theta_0^2\ \frac{\varepsilon\ \mathsf{P}_0}{\alpha} .$$

Thus, the triangle inequality gives

$$\sup_{\mathbb{T}^d_{s_*}} |\mathsf{W}_0(\phi_* - \phi_{\rm e})| \leqslant \mathsf{C}_{13}\, \sigma_0^{-(6\nu+3d+2)}\, \theta_0^3\, \epsilon_0 \;,$$

which proves the bounds on $\|u_*\|$ and $\|v_*\|$ in (2.6). Let us now prove the bound on $\partial_x u_*$ in (2.6). Set

$$\tilde u_j := \partial_x \pi_2(\phi_j - {\rm id}), \qquad U^j := \partial_x \pi_2 \phi_0 \circ \phi^j = (\mathbb{1}_d + \tilde u_0) \circ \cdots \circ (\mathbb{1}_d + \tilde u_j).$$

Then, for any $j \geqslant 0$, we have

$$\|U^j\|_{s_{i+1}} \leqslant (1 + \|\tilde u_0\|_{s_0}) \cdots (1 + \|\tilde u_j\|_{s_j}) \overset{(3.30)+(3.43)}{\leqslant} \exp\left(d^{-2} \sum_{k\geqslant 0} \sigma_k^{d-1}\, \varepsilon^{2^k} \mathsf{L}_k \right) \leqslant e^{\gamma} \;,$$

so that

$$\|U^{j+1} - U^j\|_{s_*} = \|U^j(\mathbb{1}_d + \tilde u_{j+1}) - U^j\|_{s_*} \leqslant \|U^j\|_{s_{j+1}} \|\tilde u_{j+1}\|_{s_{j+1}} \overset{(3.30)+(3.43)}{\leqslant} e^{\gamma} d^{-2} \sigma_{j+1}^{d-1}\, \varepsilon^{2^{j+1}} \mathsf{L}_{j+1},$$

which implies

$$\|U^j - \mathbb{1}_d\|_{s_*} \leqslant e^{\gamma} d^{-2} \sum_{k\geqslant 0} \sigma_k^{d-1}\, \varepsilon^{2^k} \mathsf{L}_k \leqslant \gamma\, e^{\gamma} \overset{(2.4)}{\leqslant} e\,\gamma \overset{(2.4)}{\leqslant} \frac{1}{2}$$

and then, letting $j \to \infty$, we get the estimate on $\partial_x u_*$. □

Remark 2. As it is easy to check, Theorem A holds under the milder condition $\epsilon \leqslant \epsilon_\sharp$ where

$$\begin{aligned} \epsilon_\sharp \;&:= \max\Big\{ \epsilon > 0 \;:\; \mathsf{C}_{14}\, \theta^{\frac{41}{8}}\, (s - s_*)^{-2(3\nu+2d+1)}\, \epsilon^2 \left(\log \epsilon^{-1}\right)^{2\nu} \leqslant 1 \;, \\ &\qquad \text{and} \qquad 2\mathsf{C}\, (s - s_*)^{-(6\nu+3d+2)}\, \theta^3\, \epsilon\, \exp\left(\mathsf{C}\, (s - s_*)^{-(6\nu+3d+2)}\, \theta^3\, \epsilon \right) \leqslant 1 \Big\} \;. \end{aligned}$$

Notice that $\epsilon_* < \epsilon_\sharp$.
Indeed, the condition

$$\mathsf{C}_{14}\, \theta^{\frac{41}{8}}\, (s - s_*)^{-2(3\nu+2d+1)}\, \epsilon^2 \left(\log \epsilon^{-1}\right)^{2\nu} \leqslant 1$$

guaranties the convergence of Arnold's scheme, while the condition

$$2\mathsf{C}\, (s - s_*)^{-(6\nu+3d+2)}\, \theta^3\, \epsilon\, \exp\left(\mathsf{C}\, (s - s_*)^{-(6\nu+3d+2)}\, \theta^3\, \epsilon \right) \leqslant 1$$

ensures that the torus $\mathcal{T}_{\omega,\varepsilon}$ is a Lagrangian graph (over the "angle" variables).

APPENDIX A. CONSTANTS

For convenience, we collect here the list of constants appearing in the proof of Theorem A. Recall that $\tau \geqslant d - 1 \geqslant 1$ and notice that all C_i's are greater than 1 and depend only upon d and τ.

$$\nu := \tau + 1 \;,$$

$$\mathsf{C}_0 := 4\sqrt{2} \left(\frac{3}{2}\right)^{2\nu+d} \int_{\mathbb{R}^d} \left(|y|_1^{\nu} + |y|_1^{2\nu}\right) e^{-|y|_1} dy \;, \quad \mathsf{C}_1 := 2 \left(\frac{3}{2}\right)^{\nu+d} \int_{\mathbb{R}^d} |y|_1^{\nu}\, e^{-|y|_1} dy \;,$$

$$\mathsf{C}_2 := 2^{3d} d \;, \quad \mathsf{C}_3 := \left(d^2 \mathsf{C}_1^2 + 6d\mathsf{C}_1 + \mathsf{C}_2\right) \sqrt{2} \;, \quad \mathsf{C}_4 := \max\left\{6d^2 \mathsf{C}_0,\, \mathsf{C}_3\right\} \;,$$

$$\mathsf{C}_5 := \frac{3 \cdot 2^5 d}{5} \;, \quad \mathsf{C}_6 := \max\left\{2^{2\nu} \,,\, \mathsf{C}_5\right\} \;, \quad \mathsf{C}_7 := 3d \cdot 2^{6\nu+2d+3} \sqrt{2} \max\left\{640 d^2 \,,\, \mathsf{C}_4\right\} \;,$$

$$\mathsf{C}_8 := \left(2^{-d} \mathsf{C}_6\right)^{\frac{1}{8}} \;, \quad \mathsf{C}_9 := 3 \max\left\{80 d\sqrt{2} \,,\, \mathsf{C}_4\right\} \;, \quad \mathsf{C}_{10} := \left(2^{-(4\nu+2d+1)} + 2\mathsf{C}_7\right) \mathsf{C}_9/(3d^2) \;,$$

$$\mathsf{C}_{11} := \frac{1}{2^{5\nu+3d-2}} + \frac{\mathsf{C}_7\, \mathsf{C}_9}{3 \cdot 5 \cdot 2^{\nu+2} \cdot d^2 \cdot \sqrt{2}} \;, \quad \mathsf{C}_{12} := 2^{2\nu+2d+1}\, \mathsf{C}_6^2\, \mathsf{C}_7\, \mathsf{C}_8\, \mathsf{C}_9 \;,$$

$$\mathsf{C}_{13} := \mathsf{C}_{10} + 2^{-(\nu+1)}\, \mathsf{C}_{11} \;, \quad \mathsf{C}_{14} := 2^{2(3\nu+2d+1)} \mathsf{C}_{12} \;, \quad \mathsf{C}_{15} := 18 d^3 + 70 \;,$$

$$\mathsf{C} \;:= 2^{6\tau+3d+8} \mathsf{C}_{13} \;, \quad \mathsf{C}_* := \max\left\{ (4\nu\, e^{-1})^{8\nu/3} \mathsf{C}_{14}^{2/3},\; 2\mathsf{C}_{15}\mathsf{C} \right\} \;.$$

APPENDIX B. KOLMOGOROV'S NON-DEGENERACY

Let

$$\hat{\epsilon} := 2\, e\ \mathsf{C}\ (s - s_*)^{-(6\tau+3d+8)}\ \theta^3\ \cdot \frac{\mathsf{K}\ \varepsilon \mathsf{P}}{\alpha^2}\ .$$

Since $\|\partial_x u_*\|_{s_*} \overset{(2.6)}{\leqslant} 1/2$, it follows that $\mathrm{id} + u_*$ is a diffeomorphism of $\mathbb{T}^d$. Letting

$$(\partial_x(\mathrm{id} + u_*)(x))^{-1} =: \mathbb{1}_d + A(x)\ ,$$

we have

$$\|A\|_{s_*} \leqslant 2\|\partial_x u_*\|_{s_*} \overset{(2.6)}{\leqslant} 2\hat{\epsilon} \overset{(2.4)}{\leqslant} 1\ ; \quad \|v_*\|_{s_*} \overset{(2.6)}{\leqslant} \frac{\alpha}{\mathsf{K}} \frac{\hat{\epsilon}}{2\, e} \overset{(2.4)}{\leqslant} \frac{r}{\theta} \frac{1}{4\, e \mathsf{C}_{15}} < \frac{r}{8}. \tag{B.1}$$

In [15] it is proven that the map

$$\phi(y, x) := (y_0 + v_*(x) + y + A^T y, x + u_*(x))$$

is symplectic. Then

$$H \circ \phi(y, x) = E + \omega \cdot y + Q(y, x)$$

with

$$\begin{aligned}
&E = K(y_0), \quad \langle Q_{yy}(0, \cdot)\rangle = K_{yy}(y_0) + \langle M\rangle\ ,\\
&M := \partial_y^2 \bigg(K(y_0 + v_* + y + A^T y) - \frac{1}{2} y^T K_{yy}(y_0) y \bigg)\Big|_{y=0} + \partial_y^2 (\varepsilon P \circ \phi)\Big|_{y=0}\ ,\\
&\sup_{\mathbb{T}^d_{s_*}} \|K_{yy}(y_0)^{-1} M\| \overset{(B.1)}{\leqslant} (18d^3 + 70)\hat{\epsilon}\theta \overset{(2.4)}{\leqslant} 1/2,
\end{aligned}$$

which shows that $\langle Q_{yy}(0, \cdot)\rangle$ is invertible.

APPENDIX C. REMINDERS

Classical Estimates (Cauchy, Fourier)

Lemma 4. [4] *Let $p \in \mathbb{N}$, $r, s > 0, y_0 \in \mathbb{C}^d$ and let f be a real-analytic function $D_{r,s}(y_0)$ with*

$$\|f\|_{r,s} := \sup_{D_{r,s}(y_0)} |f|.$$

Then
(i) *For any multi-index $(l, k) \in \mathbb{N}^d \times \mathbb{N}^d$ with $|l|_1 + |k|_1 \leqslant p$ and for any $0 < r' < r$, $0 < s' < s$,*[16]

$$\|\partial_y^l \partial_x^k f\|_{r',s'} \leqslant p!\ \|f\|_{r,s} (r - r')^{|l|_1} (s - s')^{|k|_1}.$$

(ii) *For any $k \in \mathbb{Z}^d$ and any $y \in D_r(y_0)$*

$$|f_k(y)| \leqslant e^{-|k|_1 s} \|f\|_{r,s}.$$

[16] As usual, $\partial_y^l := \frac{\partial^{|l|_1}}{\partial y_1^{l_1} \cdots \partial y_d^{l_d}}$, $\forall\, y \in \mathbb{R}^d$, $l \in \mathbb{Z}^d$.

Implicit Function Theorem

Lemma 5. [6] *Let $r, s > 0$, $n, m \in \mathbb{N}$, $(y_0, x_0) \in \mathbb{C}^n \times \mathbb{C}^m$ and*[17]

$$F\colon (y, x) \in D^n_r(y_0) \times D^m_s(x_0) \subset \mathbb{C}^{n+m} \mapsto F(y, x) \in \mathbb{C}^n$$

be continuous with continuous Jacobian matrix F_y. Assume that $F_y(y_0, x_0)$ is invertible with inverse $T := F_y(y_0, x_0)^{-1}$ such that

$$\sup_{D^n_r(y_0)\times D^m_s(x_0)} \|\mathbb{1}_n - TF_y(y, x)\| \leqslant c < 1 \quad \textit{and} \quad \sup_{D^m_s(x_0)} |F(y_0, \cdot)| \leqslant \frac{(1-c)r}{\|T\|}. \tag{C.1}$$

Then there exists a unique continuous function $g\colon D^m_s(x_0) \to D^n_r(y_0)$ such that the following are equivalent:

(i) $(y, x) \in D^n_r(y_0) \times D^m_s(x_0)$ *and* $F(y, x) = 0$*;*

(ii) $x \in D^m_s(x_0)$ *and* $y = g(x)$*.*

Moreover, g satisfies

$$\sup_{D^m_s(x_0)} |g - y_0| \leqslant \frac{\|T\|}{1-c} \sup_{D^m_s(x_0)} |F(y_0, \cdot)|. \tag{C.2}$$

FUNDING

L. C. has been supported by the ERC grant HamPDEs under FP7 no. 306414 and the PRIN national grant "Variational Methods in Analysis, Geometry and Physics". The authors are indebted to an anonymous referee for valuable suggestions and corrections.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. Arnol'd, V. I., Proof of a Theorem of A. N. Kolmogorov on the Invariance of Quasi-Periodic Motions under Small Perturbations of the Hamiltonian, *Russian Math. Surveys,* 1963, vol. 18, no. 5, pp. 9–36; see also: *Uspekhi Mat. Nauk,* 1963, vol. 18, no. 5, pp. 13–40.
2. Arnol'd, V. I., Kozlov, V. V., and Neĭshtadt, A. I., *Mathematical Aspects of Classical and Celestial Mechanics,* 3rd ed., Encyclopaedia Math. Sci., vol. 3, Berlin: Springer, 2006.
3. Biasco, L. and Chierchia, L., Explicit Estimates on the Measure of Primary KAM Tori, *Ann. Mat. Pura Appl. (4),* 2018, vol. 197, no. 1, pp. 261–281.
4. Celletti, A. and Chierchia, L., A Constructive Theory of Lagrangian Tori and Computer-Assisted Applications, in *Dynamics Reported. Expositions in Dynamical Systems,* C. K. R. T. Jones, U. Kirchgraber, H.-O. Walther (Eds.), Dynam. Report. Expositions Dynam. Systems (N. S.), vol. 4, Berlin: Springer, 1995, pp. 60–129.
5. Chierchia, L., A. N. Kolmogorov's 1954 Paper on Nearly-Integrable Hamiltonian Systems, *Regul. Chaotic Dyn.,* 2008, vol. 13, no. 2, pp. 130–139.
6. Chierchia, L., Kolmogorov – Arnold – Moser (KAM) Theory, in *Mathematics of Complexity and Dynamical Systems: Vol. 2,* R. A. Meyers (Ed.), New York: Springer, 2012, pp. 810–836.
7. Chierchia, L. and Gallavotti, G., Smooth Prime Integrals for Quasi-Integrable Hamiltonian Systems, *Il Nuovo Cimento B,* 1982, vol. 67, no. 2, pp. 277–295.
8. Chierchia, L. and Koudjinan, E. C., *Structure of the Kolmogorov's Set,* in preparation (2019).
9. Chierchia, L. and Procesi, M., Kolmogorov – Arnold – Moser (KAM) Theory for Finite and Infinite Dimensional Systems, in *Encyclopedia of Complexity and Systems Science Living Edition,* R. A. Meyers (Ed.), 2nd ed., New York: Springer, 2018.

[17] Here, $D^n_r(z_0)$ denotes the ball in $\mathbb{C}^n$ centred at z_0 and with radius r.

10. Delshams, A. and Gutiérrez, P., Effective Stability and KAM Theory, *J. Differential Equations,* 1996, vol. 128, no. 2, pp. 415–490.
11. Kolmogorov, A. N., Preservation of Conditionally Periodic Movements with Small Change in the Hamilton Function, in *Stochastic Behaviour in Classical and Quantum Hamiltonian Systems (Volta Memorial Conference, Como, 1977),* G. Casati, J. Ford (Eds.), Lect. Notes Phys. Monogr., vol. 93, Berlin: Springer, 1979, pp. 51–56; see also: *Dokl. Akad. Nauk SSSR (N. S.),* 1954, vol. 98, pp. 527–530.
12. Koudjinan, C. E., Quantitative KAM normal forms and sharp measure estimates, *PhD Thesis,* Rome: Università degli Studi Roma Tre, 2019.
13. Pöschel, J., Integrability of Hamiltonian Systems on Cantor Sets, *Comm. Pure Appl. Math.,* 1982, vol. 35, no. 5, pp. 653–696.
14. Neishtadt, A. I., Estimates in the Kolmogorov Theorem on Conservation of Conditionally Periodic Motions, *J. Appl. Math. Mech.,* 1981, vol. 45, no. 6, pp. 766–772; see also: *Prikl. Mat. Mekh.,* 1981, vol. 45, no. 6, pp. 1016–1025.
15. Salamon, D. A., The Kolmogorov – Arnold – Moser Theorem, *Math. Phys. Electron. J.,* 2004, vol. 10, Paper 3, 37 pp.
16. Villanueva, J., Kolmogorov Theorem Revisited, *J. Differential Equations,* 2008, vol. 244, no. 9, pp. 2251–2276.