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 \( \varepsilon \to 0 \), \( \varepsilon \) 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 \( \varepsilon \), of Lagrangian invariant tori of analytic integrable systems governed by a nearly integrable Hamiltonian

\[
H(y,x) = K(y) + \varepsilon P(y,x)
\]

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 \( \varepsilon \), 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 \( \phi^t_K \) governed by the integrable Hamiltonian \( K \), is such that the frequency \( \omega := K_y(y_0) \) is Diophantine, i.e., it satisfies, for some \( \alpha, \tau > 0 \),

\[
|\omega \cdot k| := \sum_{j=1}^d |\omega_j k_j| \geq \frac{\alpha}{|k|^\tau}, \quad \forall \ k \in \mathbb{Z}^d \setminus \{0\};
\]

*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 $T_{\omega,0}$ deforms, for positive small enough $\varepsilon$, into a Lagrangian torus $T_{\omega,\varepsilon}$ invariant for $\phi_t^H$.

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 $\varepsilon$ and the size of the small divisors appearing in the iterative scheme, the size of which may be measured by the “homogeneous Diophantine constant” $\alpha$ (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 $\varepsilon$,

$$H(y, x) = \frac{1}{2} y^2 + \varepsilon (\cos x - 1)$$ \hspace{1cm} (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. Kolmogov, 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 $\varepsilon$-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 $S := \{|y| \leq 2\sqrt{\varepsilon}\}$ 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" $S$ are such that:

$$|y_0| > 2\sqrt{\varepsilon}.$$  \hspace{1cm} (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

$$\frac{\varepsilon}{\alpha^2} < \frac{1}{4}. \hspace{1cm} (1.5)$$

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

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

Thus, one has that

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

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

$$\text{osc}(v_\varepsilon) \geq \frac{4}{1 + \sqrt{2}} \cdot \frac{\varepsilon}{\alpha}. \hspace{1cm} (1.6)$$

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

$$\frac{\varepsilon}{\alpha^2} < c \hspace{1cm} (1.7)$$

(for an $\varepsilon$ and $\alpha$ 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 $\{y = y_0 + v_\alpha(x), x \in \mathbb{T}^d\}$ of the form (compare (2.6) below)

$$\text{osc}(v_\alpha) \leq C \cdot \frac{\varepsilon}{\alpha} \hspace{1cm} (1.8)$$

(for an $\varepsilon$ and $\alpha$ independent constant $C$), which, in view of (1.6), is seen to be optimal (as far as the dependence upon $\varepsilon$ and $\alpha$ 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 $\sup_{x,x'}|U(x) - U(x')|$. 

REGULAR AND CHAOTIC DYNAMICS Vol. 24 No. 6 2019
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”

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

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:

$$\partial_y K_j(y_j) = \omega := \partial_y K(y_0), \quad \det \partial^2_y K_j(y_j) \neq 0,$$

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 $\phi_j$ symplectic, closer and closer to the identity. The persistent torus $T_{\omega, \varepsilon}$ is then obtained as the limit

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

The symplectic transformations $\phi_j$’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 $\kappa_j$, which allows him to deal with a finite number of small divisors. In view of the exponential decay of Fourier coefficients, $\kappa_j$ can be taken as

$$\det \begin{pmatrix} \partial^2_y K & \partial_y K \\ \partial_y K & 0 \end{pmatrix} \bigg|_{y = y_0} \neq 0,$$

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$^6$

$$\det \begin{pmatrix} \partial^2_y K & \partial_y K \\ \partial_y K & 0 \end{pmatrix} \bigg|_{y = y_0} \neq 0.$$

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 $\partial^2_y K$ 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 \( |x| := \max_{1 \leq j \leq n} |x_j| \) be the sup-norm.

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

• \( \pi_1 : \mathbb{C}^d \times \mathbb{C}^d \ni (y, x) \mapsto -\vec{y} \) and \( \pi_2 : \mathbb{C}^d \times \mathbb{C}^d \ni (y, x) \mapsto -\vec{x} \) are the projections on the first and second component, respectively.

• For \( \alpha > 0, \tau \geq d - 1 \geq 1 \),

\[
\Delta^\tau_{\alpha} := \left\{ \omega \in \mathbb{R}^d : |\omega \cdot k| \geq \frac{\alpha}{|k|_1}, \quad \forall 0 \neq k \in \mathbb{Z}^d \right\} \quad (2.1)
\]

is the set of \((\alpha, \tau)\)-Diophantine numbers in \( \mathbb{R}^d \).

• For \( r, s > 0, y_0 \in \mathbb{C}^d \), we denote:

\[
\mathbb{T}^d_s := \left\{ x \in \mathbb{C}^d : |\text{Im } x| < s \right\} / 2\pi \mathbb{Z}^d,
\]

\[
B_r(y_0) := \left\{ y \in \mathbb{R}^d : |y - y_0| < r \right\}, \quad (y_0 \in \mathbb{R}^d),
\]

\[
D_r(y_0) := \left\{ y \in \mathbb{C}^d : |y - y_0| < r \right\}, \quad D_{r,s}(y_0) := D_r(y_0) \times \mathbb{T}^d_s.
\]

• If \( I_d := \text{diag}(1) \) is the unit \((d \times d)\) matrix, we denote the standard symplectic matrix by

\[
J := \begin{pmatrix}
0 & -I_d \\
I_d & 0
\end{pmatrix}.
\]

• 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:

\[
\| \cdot \|_{r,y_0} := \sup_{D_r(y_0)} | \cdot |, \quad \| \cdot \|_s := \sup_{\mathbb{T}^d_s} | \cdot |.
\]

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

• 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

\[
\| \mathcal{L} \| := \sup_{x \in V_1 \setminus \{0\}} \frac{\| \mathcal{L} x \|_2}{\| x \|_1}, \quad \text{so that } \| \mathcal{L} x \|_2 \leq \| \mathcal{L} \| \| x \|_1 \text{ for any } x \in V_1.
\]
Given $\omega \in \mathbb{R}^d$, the directional derivative of a $C^1$ function $f$ with respect to $\omega$ 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
\[ f = \sum_{k \in \mathbb{Z}^d} f_k e^{ik \cdot x}, \quad f_k := \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} f(x) e^{-ik \cdot x} \, dx, \]
(where, as usual, $e := \exp(1)$ denotes the Neper number and $i$ the imaginary unit). We also set:
\[ \langle f \rangle := f_0 = \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} f(x) \, dx, \quad (p_N f)(x) := \sum_{|k|_1 \leq N} f_k e^{ik \cdot x}, N > 0, \]
$p_N$ being the Fourier projection onto the Fourier modes with $|k|_1 \leq N$; notice that $\langle \cdot \rangle = p_0(\cdot)$.

We are ready to formulate a quantitative version of Arnold’s theorem.

**Theorem A.** Let $d \geq 2; \tau \geq d - 1; \alpha, r, \varepsilon > 0; 0 < s_* < s \leq 1; y_0 \in \mathbb{R}^d; K, P \in \mathcal{A}_{r,s}(y_0); H := K + \varepsilon P$. Assume that
\[ \left\{ \begin{array}{l}
\omega := \partial_y K(y_0) \in \Delta_\alpha^\tau, \\
\det(\partial^2_y K(y_0)) \neq 0.
\end{array} \right. \tag{2.2} \]
Define:
\[ T := \partial^2_y K(y_0)^{-1}, P := \|P\|_{r,s,y_0}, K := \|\partial^2_y K\|_{r,y_0}, T := \|T\|, \theta := TK, \]
and denote by $\varepsilon$ the rescaled smallness parameter:
\[ \varepsilon := \frac{KP}{\alpha^r}. \tag{2.3} \]
There exist constants $1 < C < C_*$ depending only on $d$ and $\tau$, such that, if $a := 6\tau + 3d + 8$ and
\[ \alpha \leq \frac{r}{T} \quad \text{and} \quad \varepsilon \leq \varepsilon_* := \frac{(s - s_*)a}{C_\ast \theta^4}, \tag{2.4} \]
then there exists a real-analytic embedding
\[ \phi_* : x \in \mathbb{T}^d_{s_*} \mapsto \phi_*(x) := \phi_e(y_0, x) + (v_*(x), u_*(x)) \in D_{r,s}(y_0), \]
where $\phi_e$ is the trivial embedding
\[ \phi_e : x \in \mathbb{T}^d \mapsto (y_0, x), \]
such that the $d$–torus
\[ T_{\omega,\varepsilon} := \phi_* \left( \mathbb{T}^d \right) \tag{2.5} \]
is a Lagrangian torus satisfying
\[ \phi^*_H \circ \phi_*(x) = \phi_*(x + \omega t), \quad \forall \ x \in \mathbb{T}^d_{s_*}, \forall \ t \in \mathbb{R}. \]

To avoid introducing too many symbols, we use capital straight style for positive constants ($P, K, T, C, \ldots$), while usually capital normal style is used for functions or matrices ($K, P, H, T, \ldots$).
Lemma 1. \[ \text{Let } \theta := TK \geq T\|K_{yy}(y_0)\| \geq \|T\|\|K_{yy}(y_0)\| = \|T\|\|T^{-1}\| \geq 1. \] (2.7)

Remarks and addenda.

(i) \( \theta \) is a measure of the local “torsion” and is a number greater than or equal to one: \[ \theta := TK \geq T\|K_{yy}(y_0)\| \geq \|T\|\|K_{yy}(y_0)\| = \|T\|\|T^{-1}\| \geq 1. \] (2.7)

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

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

(iv) The torus \( T_{\omega,\epsilon} \) 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 \( \phi \) close to \( \phi_0 \), for which
\[ H \circ \phi(y, x) = E + \omega \cdot y + Q(y, x) \] such that \( \det(Q_{yy}(0, \cdot)) \neq 0; \) for details, see Appendix B.

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

(vi) The dependence of the invariant torus \( T_{\omega,\epsilon} \) on \( \epsilon \) 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 \( T_{\omega,\epsilon} \) 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:

\[ \nu := \tau + 1, \quad C_0 := 4\sqrt{2} \left( \frac{3}{2} \right)^{2\nu+d} \int_{\mathbb{R}^d} \left( |y|_1^{\nu} + |y|_1^{2\nu} \right) e^{-|y|_1^2} dy, \]

\[ C_1 := 2 \left( \frac{3}{2} \right)^{\nu+d} \int_{\mathbb{R}^d} |y|_1^{\nu} e^{-|y|_1^2} dy, \]

\[ C_2 := 2^{3d} d, \quad C_3 := (d^2C_1^2 + 6dC_1 + C_2) \sqrt{2}, \quad C_4 := \max \{ 6d^2C_0, C_3 \}. \]

Lemma 1. Let\(^{8)} \; r > 0, \; 0 < 2\sigma < s \leq 1, \; y \in \mathbb{R}^d, \; K, P \in \mathcal{A}_{r,s}(y) \) and consider the Hamiltonian parametrised by \( \epsilon > 0 \)
\[ H(y, x; \epsilon) := K(y) + \epsilon P(y, x). \]

Assume that
\[ \det K_{yy}(y) \neq 0, \quad \omega := K_y(y) \in \Delta_\alpha^r, \]

\(^{8)}K \) and \( \rho \) 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
\[
\|K_{yy}\|_{r,y} \leq K, \quad \|T\| \leq T, \quad \|P\|_{r,s,y} \leq P,
\]  
(3.1)
where $T := K_{yy}(y)^{-1}$.

Now let $\lambda, \bar{r}, \tilde{r}$ be positive numbers such that:
\[
\lambda \geq \log \left( \frac{\sigma^{2\nu+d} \alpha^2}{\varepsilon PK} \right), \quad \bar{r} \leq \frac{5}{24dTK}, \quad \tilde{r} \leq \min \left\{ \bar{r}, \frac{\alpha}{2dK\kappa^{r+1}} \right\},
\]  
(3.2)
where
\[
\kappa := \frac{4\lambda}{\sigma}.
\]
Finally, define
\[
L := P \max \left\{ \frac{40dT^2K}{r^2} \sigma^{-(\nu+d)}, \frac{C_4}{\sqrt{2}} \max \left\{ 1, \frac{\alpha}{TK} \right\} \frac{K}{\alpha^2} \sigma^{-2(\nu+d)} \right\}, \quad \bar{s} := s - \frac{2}{3} \sigma, \quad s' := s - \sigma.
\]

Then, if
\[
\varepsilon L \leq \frac{\sigma}{3},
\]  
(3.3)
there exist $y' \in \mathbb{R}^d$ and a symplectic change of coordinates
\[
\phi' = \text{id} + \varepsilon \phi : D_{\bar{r}/2,s'}(y') \to D_{2\bar{r}/3,s}(y),
\]  
(3.4)
such that
\[
\left\{ \begin{array}{l}
H \circ \phi' = H' = K' + \varepsilon^2 P', \\
\partial_{y'} K'(y') = \omega, \quad \det \partial_{y'}^2 K'(y') \neq 0,
\end{array} \right.
\]  
(3.5)
where
\[
K' := K + \varepsilon \tilde{K} := K + \varepsilon \langle P(y', \cdot) \rangle.
\]
Moreover, letting
\[
(\partial_{y'}^2 K'(y'))^{-1} = T + \varepsilon \tilde{T},
\]
the following estimates hold:
\[
\|\partial_{y'}^2 \tilde{K}\|_{r/2,y} \leq KL, \quad |y' - y| \leq \frac{8\varepsilon TP}{r}, \quad \|\tilde{T}\| \leq TL,
\]  
(3.6)
\[
\max\{\|\partial_{y} \pi_2 \phi\|_{s'}, \|W \phi\|_{r/2,s',y'}\} \leq d^{-2} \sigma^{d-1} L, \quad \|P'\|_{r/2,s',y'} \leq LP,
\]  
(3.7)
where
\[
W := \begin{pmatrix}
\max\{\frac{K}{\alpha}, \frac{1}{r}\} & 1_d & 0 \\
0 & 1_d
\end{pmatrix}.
\]

Observe that
\[
\sigma^{-2(\nu+d)} \varepsilon PK / \alpha^2 \leq (\sqrt{2}/C_4) \varepsilon L,
\]
so that (3.3) implies
\[
\frac{\varepsilon PK}{\alpha^2} < \frac{\sigma^{2\nu+d}}{e},
\]
which, in particular, implies that $\lambda > 1$ and $\kappa > 4$. 

REGULAR AND CHAOTIC DYNAMICS Vol. 24 No. 6 2019
Proof.

Step 1: Construction of Arnold’s transformation

We seek a near-identity symplectic transformation

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

with $D_{r_1,s_1}(y') \subset D_{r,s}(y)$, generated by a generating function of the form

$$y' = y + \varepsilon g_x(y',x)$$

and

$$x' = x + \varepsilon g_y(y',x),$$

such that

$$H' := H \circ \phi' = K' + \varepsilon^2 P',$$

$$\left(\partial_y K'(y') = \omega, \quad \det \partial_{y'}^2 K'(y') \neq 0.\right)$$

By Taylor’s formula, we get

$$H(y' + \varepsilon g_x(y',x), x) = K(y') + \varepsilon K'(y') + \varepsilon \left[K'(y') \cdot g_x + P_\kappa P(y',\cdot) - \tilde{K}(y')\right]$$

$$+ \varepsilon^2 \left(P^{(1)} + P^{(2)} + P^{(3)}\right)(y',x)$$

$$= K'(y') + \varepsilon \left[K'(y') \cdot g_x + P_\kappa P(y',\cdot) - \tilde{K}(y')\right] + \varepsilon^2 P_+(y',x),$$

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

$$\left\{\begin{array}{l}
P_+ := P^{(1)} + P^{(2)} + P^{(3)}

P^{(1)} := \frac{1}{\varepsilon^2} \left[K(y' + \varepsilon g_x) - K(y') - \varepsilon K_y(y') \cdot g_x\right] = \int_0^1 (1-t)K_{yy}(\varepsilon t g_x) \cdot g_x \cdot g_x dt

P^{(2)} := \frac{1}{\varepsilon} \left[P(y' + \varepsilon g_x, x) - P(y',x)\right] = \int_0^1 P_y(y' + \varepsilon t g_x, x) \cdot g_x dt

P^{(3)} := \frac{1}{\varepsilon} \left[P(y',x) - P_\kappa P(y',\cdot)\right] = \frac{1}{\varepsilon} \sum_{|n|_1 > \kappa} P_n(y') e^{in}.\end{array}\right.\tag{3.11}$$

By the non-degeneracy condition $\det K_{yy}(y) \neq 0$, for $\varepsilon$ small enough (to be made precise below), $\det \partial_{y'}^2 K'(y') \neq 0$ and, therefore, by the standard inverse function theorem (see, e.g., Lemma 5), there exists a unique $y' \in D_{r,y}$ 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 $K_y(y') \cdot g_x + P_\kappa P(y',\cdot) - \tilde{K}(y')$ vanishes; such a $g$ is indeed given by

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

provided that

$$K_y(y') \cdot n \neq 0, \quad \forall 0 < |n|_1 \leq \kappa, \quad \forall y' \in D_{r_1}(y') \quad (\subset D_{r}(y)).$$

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

$$K_y(y') \cdot n \neq 0, \quad \forall 0 < |n|_1 \leq \kappa, \quad \forall y' \in D_{\tilde{r}}(y).$$

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 $T^4$ and that $p_N$ is the Fourier projection onto modes with $|k|_1 \leq N$. 

REGULAR AND CHAOTIC DYNAMICS Vol. 24 No. 6 2019
The last step is to invert the function \( x \mapsto x + \varepsilon g_x(y', x) \) in order to define \( P' \). By the inverse function theorem, for \( \varepsilon \) small enough, the map \( x \mapsto x + \varepsilon g_x(y', x) \) admits a real-analytic inverse of the form
\[
\varphi_{\varepsilon}(y', x') = x' + \varepsilon \widetilde{\varphi}_{\varepsilon}(y', x'),
\]
so that Arnold’s symplectic transformation is given by
\[
\phi': (y', x') \mapsto \begin{cases} y = y' + \varepsilon g_x(y', \varphi_{\varepsilon}(y', x')) \\ x = \varphi_{\varepsilon}(y', x') = x' + \varepsilon \widetilde{\varphi}_{\varepsilon}(y', x'). \end{cases}
\]
Hence, (3.9) holds with
\[
P'(y', x') := P_+(y', \varphi_{\varepsilon}(y', x')).
\]

**Step 2: Quantitative estimates**

First of all, notice that from the definitions of \( \bar{r} \) and \( \tilde{r} \) it follows that
\[
\bar{r} \leq \tilde{r} \leq \frac{5r}{24d} < \frac{r}{2}.
\]
We begin by extending the “Diophantine condition w.r.t. \( K_y \)” uniformly to \( D_r(y) \) up to order \( \kappa \).
Indeed, by the mean value inequality and \( K_y(y) = \omega \in \Delta^\tau \), we get, for any \( 0 < |n|_1 \leq \kappa \) and any \( y' \in D_r(y) \),
\[
|K_y(y') \cdot n| = |\omega \cdot n + (K_y(y') - K_y(y)) \cdot n| \geq q |\omega \cdot n| \left( 1 - \frac{\|K_{yy}\|_{r, y}}{|\omega \cdot n|} |n|_1 \bar{r} \right)
\]
\[
\geq q |n|_1 \left( 1 - \frac{dK}{\alpha |n|^{1+\tau}} \right) \geq q |n|_1 \left( 1 - \frac{\kappa^{1+\tau}}{2n} \right) = \frac{\alpha}{2n},
\]
so that, by Fourier estimates (Lemma 4(ii)), we have
\[
\|g_x\|_{r, \bar{r}, y} = \sup_{D_{\bar{r}, \bar{r}}(y)} \left| \sum_{0 < |n|_1 \leq \kappa} \frac{n P_n(y')}{K_y(y')} \cdot n e^{\imath n \cdot x} \right| \leq \sum_{0 < |n|_1 \leq \kappa} \frac{\|P_n\|_{r, \bar{r}, y} \|n\|_1}{|K_y(y') \cdot n|} e^{(s-\frac{2}{3})|n|_1}
\]
\[
\leq \sum_{0 < |n|_1 \leq \kappa} \sum_{\alpha \in \mathbb{Z}^d} \frac{2}{\alpha} e^{-s|\alpha|} \frac{2}{\alpha} e^{-s|\alpha|} = C_\alpha \frac{\alpha}{\sigma} e^{-(\nu+d)},
\]
\[
\|\partial_y g\|_{r, \bar{r}, y} = \sup_{D_{\bar{r}, \bar{r}}(y)} \left| \sum_{0 < |n|_1 \leq \kappa} \left( \partial_y P_n(y') \cdot n - P_n(y') \right) e^{\imath n \cdot x} \right| \leq \sum_{0 < |n|_1 \leq \kappa} \frac{\|P_n\|_{r, \bar{r}, y} \|n\|_1}{|K_y(y') \cdot n|} e^{(s-\frac{2}{3})|n|_1}
\]
\[
\leq \sum_{0 < |n|_1 \leq \kappa} \left( \frac{\|P_n\|_{r, \bar{r}, y} \|n\|_1}{|K_y(y') \cdot n|} + \frac{\|K_{yy}\|_{r, y} \|n\|_1}{|K_y(y') \cdot n|} \right) e^{(s-\frac{2}{3})|n|_1}
\]
\[
\leq \sum_{0 < |n|_1 \leq \kappa} \left( \frac{4}{\alpha^2 r} e^{-s|n|_1} \frac{2|n|_1^2 \alpha + r K}{\alpha} e^{-s|n|_1} K |n|_1 \left( \frac{2|n|_1}{\alpha} \right)^2 \right) e^{(s-\frac{2}{3})|n|_1}
\]
\[
\leq \max \left\{ \alpha, r K \right\} \frac{4}{\alpha^2 r} \sum_{0 < |n|_1 \leq \kappa} \left( |n|_1^2 + |n|_1^{2\tau+1} \right) e^{-\frac{2}{3}|n|_1}
\]
\[
\leq \max \left\{ 1, \frac{\alpha}{r K} \right\} \frac{4}{\alpha^2 r} \int_{\mathbb{R}^d} \left| y_1^2 + |y_1|^{2\tau+1} \right| e^{-\frac{2}{3}|y|_1} dy
\]
Next, we prove the existence and uniqueness of $\eta$. Analogously, 

$$\frac{\partial^2 g}{\partial x^2} \mid_{r,s,y} \leq \frac{C_0}{\sqrt{2}} \max \left\{ 1, \frac{\alpha}{rK} \right\} \frac{PK}{\alpha^2} \sigma^{-(2\nu+d+1)} \leq \overline{\Gamma},$$

where 

$$\overline{\Gamma} := 6 \frac{C_0}{\sqrt{2}} \max \left\{ 1, \frac{\alpha}{rK} \right\} \frac{PK}{\alpha^2} \sigma^{-(2\nu+d+1)}.$$

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

$$\frac{\partial^3 g}{\partial x^2} \mid_{r,s,y} \leq \frac{6C_0}{\sqrt{2}} \max \left\{ 1, \frac{\alpha}{rK} \right\} \frac{PK}{\alpha^2} \sigma^{-(2\nu+d+1)} = \Gamma,$$

where 

$$s'' := s - \frac{\alpha}{6} \sigma$$

and 

$$\| \frac{\partial^3 g}{\partial x^2} \mid_{r,s''} \|_{x} := \sup_{D_{r,s''}(y)} \max \{ | \frac{\partial^3 g}{\partial x^2} \mid_{r,s''} \|_{x} : i, j, k = 1, \ldots, d \}.$$

Also, 

$$\| \frac{\partial^2 K}{\partial y^2} \|_{r/2,y} = \| \langle P_y \rangle \|_{r/2,y} \leq \| P_y \|_{r/2,y} \leq \frac{P}{r - \frac{r}{2}} \leq \frac{2P}{r},$$

$$\| \frac{\partial^2 K}{\partial y^2} \|_{r/2,y} = \| \langle P_{yy} \rangle \|_{r/2,y} \leq \| P_{yy} \|_{r/2,y} \leq \frac{P}{(r - \frac{r}{2})^2} \leq \frac{4P}{r^2} \leq \frac{KL}{2}.$$

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

$$F : D_r(y) \times U_\varepsilon \rightarrow \mathbb{C}^d$$

$$(y, \eta) \mapsto K_y(y) + \eta \tilde{K}_y(y) - K_y(y).$$

Then

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

$$\| \mathbb{I} - T \mathbb{F}_y(y, \eta) \| \leq \| \mathbb{I} - T K_{yy} \| + |\eta| \| T \| \| \frac{\partial^2 \tilde{K}}{\partial y^2} \|_{r/2,y}$$

$$\leq \| \mathbb{I} - T K_{yy} \| + |\eta| \| T \| \frac{4P}{r^2} \leq \frac{TK}{r - \frac{r}{2}} + 8T \varepsilon \frac{P}{r^2} \leq \frac{dTK}{r} + \frac{8TP}{r^2}$$

$$\leq 2dTK \frac{r}{r} + \frac{1}{2} \varepsilon L \leq \frac{5}{12} + \frac{\sigma}{6} \leq \frac{5}{12} + \frac{1}{12} = \frac{1}{2}.$$
Recalling \( \sigma \leq \frac{1}{2} \), we have

\[
2\|T\|\|F(y, \cdot)\|_{2\varepsilon, 0} = 2\|T\| \sup_{\tilde{u}_i} |\eta \tilde{K}_y(y)| \leq 2T \frac{4\varepsilon P}{r} \leq \frac{5 \cdot 2^\nu + d}{8d} \frac{r}{\text{TK}} \sigma^{\nu + d} \varepsilon L \\
= 3 \cdot 2^d (2\sigma)^\nu \tilde{r}^{\sigma} \delta \varepsilon L \leq 3 \cdot 2^d \tilde{r}^{\sigma} \delta \varepsilon L \leq \frac{3}{2} \tilde{r}^{\sigma} \delta \varepsilon L.
\]

Therefore, we can apply the inverse function theorem (Lemma 5). Hence, there exists a function \( D_{\tilde{r}}(y) \) satisfying \( 0 = F(y, \varepsilon) = \partial y K'(y) - \omega \), i.e., the second part of (3.9). Moreover,

\[
|y' - y| \leq 2\|T\|\|F(y, \cdot)\|_{2\varepsilon, 0} \leq \frac{8\varepsilon TP}{r} \leq 3 \cdot 2^d \tilde{r}^{\sigma} \delta \varepsilon L \leq \frac{\tilde{r}}{2},
\]

so that

\[
D_{\tilde{r}}(y') \subset D_r(y).
\]

Next, we prove that \( \partial^2_y K'(y') \) is invertible. Indeed, by Taylor’s formula, we have

\[
\partial^2_y K'(y') = K_{yy}(y) + \int_0^1 K_{yyyy}(y + t\varepsilon \tilde{y}) \cdot \varepsilon \tilde{y} dt + \varepsilon \tilde{K}_{yy}(y')
\]

\[
= T^{-1} \left( 1_d + \varepsilon T \left( \int_0^1 K_{yyyy}(y + t\varepsilon \tilde{y}) \cdot \tilde{y} dt + \tilde{K}_{yy}(y') \right) \right)
\]

\[
= T^{-1}(1_d + \varepsilon A),
\]

and, by Cauchy’s estimate,

\[
\varepsilon \|A\| \leq \|T\| \left( d \|K_{yyyy}\|_{r/2, y} \varepsilon |y' - y| + \varepsilon \|\partial^2_y \tilde{K}\|_{r/2, y} \right)
\]

\[
\leq \|T\| \left( \frac{d \|K_{yyyy}\|_{r, y}}{r - \frac{r}{2}} \varepsilon |y' - y| + \varepsilon \|\tilde{K}_{yy}\|_{r/2, y} \right)
\]

\[
\leq T \left( \frac{2dK}{r} \frac{8\varepsilon TP}{r} + \frac{4\varepsilon P}{r^2} \right) \leq \frac{4\varepsilon TP}{r^2} (4dTK + 1)
\]

\[
\leq \frac{20d\varepsilon T^2 KP}{r^2} \leq \frac{1}{2} \varepsilon L \leq \frac{\sigma}{6} \leq \frac{1}{2}.
\]

Hence, \( \partial^2_y K'(y') \) is invertible with

\[
\partial^2_y K'(y')^{-1} = (1_d + \varepsilon A)^{-1} T = T + \sum_{k \geq 1} (-\varepsilon)^k A^k T = T + \varepsilon \tilde{T},
\]

and

\[
\varepsilon \|\tilde{T}\| \leq \varepsilon \frac{\|A\|}{1 - \varepsilon \|A\|} \|T\| \leq 2\varepsilon \|A\|\|T\| \leq \varepsilon LT \leq 2 \frac{\sigma}{6} T = T \frac{\sigma}{3}.
\]
and thus
\[ \| P^{(1)} \|_{r, \bar{y}, y} \leq d^2 \| K_{yy} \|_{r, y} \| g_x \|_{r, \bar{y}, y}^2 \leq d^2 K \left( \frac{P}{\alpha} \sigma^{-(\nu + d)} \right)^2 = d^2 C_1 \frac{KP^2}{\alpha^2} \sigma^{-2(\nu + d)}, \]
\[ \| P^{(2)} \|_{r, \bar{y}, y} \leq d \| P_y \|_{r, \bar{y}, y} \| g_x \|_{r, \bar{y}, y} \leq \frac{6P}{r} C_1 \frac{P}{\alpha} \sigma^{-(\nu + d)} = 6d C_1 \frac{P^2}{\alpha r} \sigma^{-2(\nu + d)}, \]
and by Fourier estimates (Lemma 4(ii)), we have
\[ \varepsilon \| P^{(3)} \|_{r, \bar{y}, y} \leq \sum_{|n|_1 > \kappa} \| P_n \|_{r, \bar{y}} e^{\left(s - \frac{\varepsilon}{2}\right)|n|_1} \leq P \sum_{|n|_1 > \kappa} e^{-\frac{\sigma|n|_1}{4}} \leq P e^{-\frac{\sigma}{4} \sum_{|n|_1 > 0} e^{-\frac{\sigma|n|_1}{4}}} \leq P e^{-\frac{\sigma}{4} \sum_{|n|_1 > \kappa} e^{-\frac{\sigma|n|_1}{4}}} \leq P e^{-\frac{\sigma}{4} \sum_{|n|_1 > \kappa} e^{-\frac{\sigma|n|_1}{4}}} \leq \varepsilon \| P \|_{r, \bar{y}, y} e^{-\lambda} \leq C_2 \sigma^{-d} \sigma^{-2(\nu + d) \varepsilon \| PK \|_{r, \bar{y}, y} e^{-\lambda} \leq C_2 \frac{P^2}{\alpha^2} \sigma^{-2(\nu + d)}. \]

Hence,
\[ \| P^+ \|_{r, \bar{y}, y} \leq \| P^{(1)} \|_{r, \bar{y}, y} + \| P^{(2)} \|_{r, \bar{y}, y} + \| P^{(3)} \|_{r, \bar{y}, y} \leq d^2 C_1 \frac{KP^2}{\alpha^2} \sigma^{-2(\nu + d)} + 6d C_1 \frac{P^2}{\alpha r} \sigma^{-\nu + d} + C_2 \varepsilon \| PK \|_{r, \bar{y}, y} e^{-\lambda} \leq C_3 \frac{KP^2}{\alpha^2} \sigma^{-2(\nu + d)} \]
\[ = \left( d^2 C_1^2 r K + 6d C_1 \alpha \sigma^{\nu + d} + C_2 r K \right) \frac{P^2}{\alpha^2 r} \sigma^{-2(\nu + d)} \]
\[ \leq \left( d^2 C_1^2 + 6d C_1 + C_2 \right) \max \left\{ \alpha, r K \right\} \frac{P^2}{\alpha^2 r} \sigma^{-2(\nu + d)} \]
\[ \leq \frac{C_3}{\sqrt{2}} \max \left\{ 1, \frac{\alpha}{r K} \right\} \frac{P^2 K}{\alpha^2} \sigma^{-2(\nu + d)} \leq \varepsilon \| P \|_{r, \bar{y}, y} e^{-\lambda} \leq \Gamma. \]

Finally, we prove that, given \( y' \in D_r(y) \), the function \( \psi_{\varepsilon}(x) = x + \varepsilon g_{y'}(y', x) \) has an analytic inverse\(^{11} \). Consider the Banach space
\[ B := \left\{ u \in C^1(T^d_y, C^d) : \| u \|_{s', 1} := \max \left\{ \| u \|_{s'} \right\} \right\}. \]
For any \( u \in B \) and any \( x' \in T^d_y \), we have \( \text{Im} \left( x' + \varepsilon u(x') \right) \leq s' + \varepsilon \| u \|_{s'} \leq s' + \varepsilon \| u \|_{s'} \leq s' + \sigma/6 = s'' \). Hence, the functional \( f : B \ni u \mapsto -g_{y'}(y', \text{id} + \varepsilon u) \) is well-defined and smooth. Moreover, for any \( u \in B \),
\[ \| f(u) \|_{s'} \leq \| g_{y'} \|_{w''} \leq \Gamma, \]
\[ \| \partial_x(f(u)) \|_{s'} \leq \| g_{y'} \|_{w''} \leq \| g_{y'} \|_{w''} \| \varepsilon \| \| \partial_x u \|_{s'} \leq \Gamma : \varepsilon \| \varepsilon \| \leq \Gamma : \varepsilon \| \varepsilon \| \leq \Gamma : \varepsilon \| \varepsilon \| \leq \frac{\sigma}{6} < \Gamma. \]

\(^{11}\)Observe that \( \psi_{\varepsilon}(\text{id} + \varepsilon u) = \text{id} \) is equivalent to \( u = -g_{y'}(y', \text{id} + \varepsilon u) \), i.e., \( u \) is a fixed point of the map \( u \mapsto -g_{y'}(y', \text{id} + \varepsilon u) \).
Thus, $f: \mathcal{B} \to \mathcal{B}$. Furthermore, for any $u_1, u_2 \in \mathcal{B}$,
\[
\|f(u_1) - f(u_2)\|_{s',1} \leq (1 + d^2 \varepsilon \Gamma) \varepsilon \cdot \|u_1 - u_2\|_{s',1} \leq \frac{2\sigma}{3d^2} \cdot \|u_1 - u_2\|_{s',1} < \frac{1}{2} \|u_1 - u_2\|_{s',1}.
\]
Hence, $f$ is a contraction. Therefore, by the Banach–Caccioppoli fixed-point theorem, $f$ has a unique fixed point $\bar{\varphi}_0 \in \mathcal{B}$; $\bar{\varphi}_0$ is obtained as the uniform limit $f^n(0)$ (as $0 \in \mathcal{B}$). Thus, as $f^0 = f$
is real-analytic on $D_f(y) \times D_y$, by Weierstrass’s theorem on the uniform convergence of analytic functions, $\bar{\varphi}_0$ is real-analytic on $D_{f,s'}(y)$. The rest of the claims on $\phi'$ and $P'$ are then obvious. □

3.2. Arnold’s Scheme: Iteration

Let $d$, $\tau$, $H$, $K$, $P$, $T$, $\varepsilon$, $\alpha$, $r$, $s$, $s_0$, $P$, $K$, $T$, $\theta$, $\epsilon$ 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\(^{12}\).

\[
\begin{align*}
\epsilon_0 &= \epsilon, \quad \theta_0 := \theta, \quad r_0 := r, \quad T_0 := T, \quad K_0 := K, \quad P_0 := P, \\
\sigma_0 &= \frac{(s - s_0)/2}{2} - \frac{1}{2}, \quad \lambda_0 := \log \epsilon^{-1}, \quad \kappa_0 := 4\sigma_0^{-1}\lambda_0, \\
C_5 &= \frac{3}{2} \cdot \frac{2^d}{5}, \quad C_6 := \max \left\{ 2^{2\nu}, C_5 \right\}, \quad C_7 := 3d \cdot 2^{6\nu+2d+3} \sqrt{2} \max \left\{ 640d^2, C_4 \right\}, \\
C_8 &= \left( 2^{-d} C_6 \right)^{\frac{1}{2}}, \quad C_9 := 3 \max \left\{ 80d\sqrt{2}, C_4 \right\}, \\
\lambda_* &= C_7 \sigma_0 \left( 1 + \frac{12\nu}{\nu + 2d} \right) \lambda_0^{2\nu}, \quad \theta_* := 2^{2\nu+2d+1} C_6^2 \theta^2, \\
\hat{\epsilon}_0 &= C_9 \sigma_0^{-1} - 2d - 1 \epsilon_0 \theta_0, \quad P_1 := \frac{\hat{\epsilon}_0 P_0}{\varepsilon}.
\end{align*}
\]

We also set, for $j \geq 0$,
\[
\begin{align*}
\sigma_j &= \frac{\sigma_0}{2^j}, \quad s_{j+1} := s_j - \tau_j = s_j + \frac{\sigma_0}{2^j}, \quad \bar{s}_j := s_j - \frac{2\sigma_j}{3}, \quad \kappa_j := 4j \kappa_0, \\
K_{j+1} &= K_0 \prod_{k=0}^j \left( 1 + \frac{\sigma_k}{3} \right) \leq K_0 \left( 1 + \frac{2\sigma_j}{3} \right) \leq K_0 \sqrt{2}, \quad T_{j+1} := T_0 \prod_{k=0}^j \left( 1 + \frac{\sigma_k}{3} \right) \leq T_0 \sqrt{2}, \\
r_{j+1} &= \frac{1}{2} \min \left\{ \frac{\alpha}{2d^{2\nu} K_0 \kappa_j}, \frac{5}{48d \theta_0} \right\}, \quad W_j := \text{diag} \left( \max \left\{ \frac{K_j}{\alpha}, \frac{1}{r_j} \right\}, \mathbb{1}_d, \mathbb{1}_d \right), \\
L_j &= P_1 \max \left\{ \frac{80d\sqrt{2} T_0 \theta_0 \sigma_j^{-(\nu + d)}}{r_j}, C_4 \max \left\{ 1, \frac{\alpha}{r_j K_j} \right\} \frac{K_0}{\alpha^2 \sigma_j^{-(\nu + d)}} \right\}.
\end{align*}
\]

Observe that
\[
W_0 = \text{diag} \left( K_0 \alpha^{-1} \mathbb{1}_d, \mathbb{1}_d \right), \quad s_j \downarrow s_0, \quad r_j \downarrow 0, \quad \epsilon_0 \leq \hat{\epsilon}_0.
\]

Note also that, since $\hat{\epsilon}_0$ is proportional to $\varepsilon$, $P_1$ is independent of $\varepsilon$.

3.2.1. First step

**Lemma 2.** Assume
\[
\alpha \leq \frac{r_0}{T_0} \quad \text{and} \quad \hat{\epsilon}_0 \leq 1.
\]  

\(^{12}\)Recall the definitions of $\nu$ 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
\[
\phi_0 : D_{r_1,s_1}(y_1) \to D_{r_0,s_0}(y_0),
\]
such that, for \( H_1 := H_0 \circ \phi_0 \), we have
\[
\begin{cases}
H_1 =: K_1 + \varepsilon^2 P_1, \\
\partial_{y_1} K_1(y_1) = \omega, \quad \det \partial_{y_1}^2 K_1(y_1) \neq 0
\end{cases}
\]
and
\[
\begin{align*}
|y_1 - y_0| &\leq \frac{8 \varepsilon T_0 P_0}{r_0}, \\
\|K_1\|_{r_1/4,y_1} &\leq K_1, \quad \|T_1\| \leq T_1, \quad T_1 := \partial_{y_1}^2 K_1(y_1)^{-1}, \\
\varepsilon^2 \|P_1\|_{r_1,s_1,y_1} &\leq \varepsilon^2 P_1, \\
\max \{ \|W_0(\phi_0 - \text{id})\|_{r_1,s_1,y_1}, \|\partial_{\theta_2}(\phi_0 - \text{id})\|_{s_1} \} &\leq d^{-2} \sigma_0^{d-1} \varepsilon L_0.
\end{align*}
\]

Proof. Since
\[
\kappa_0 \stackrel{(3.24)}{=} 4 \sigma_0^{-1} \geq 8
\]
and
\[
\frac{\alpha}{2d \sqrt{2K_0 \kappa_0^{\nu}}} \leq \frac{1}{2d \cdot 8^{\nu} \sqrt{2K_0}} \frac{r_0}{T_0} < \frac{5}{48d} \frac{r_0}{\theta_0},
\]
we get
\[
r_1 = \frac{1}{2} \min \left\{ \frac{\alpha}{2d \sqrt{2K_0 \kappa_0^{\nu}}}, \frac{5}{48d} \frac{r_0}{\theta_0} \right\} = \frac{\alpha}{4d \sqrt{2K_0 \kappa_0^{\nu}}}. (3.32)
\]
Thus,
\[
\varepsilon L_0 (3 \sigma_0^{-1}) \leq 3 \varepsilon P_0 \max \left\{ \frac{80d \sqrt{2} T_0}{r_0^2} \theta_0 \sigma_0^{-(\nu+d)} \cdot C_4 \max \left\{ 1, \frac{\alpha}{r_0 K_0} \right\} \right\} \frac{K_0}{\alpha^2} \sigma_0^{-2(\nu+d)} \leq 3 \max \left\{ \frac{80d \sqrt{2}}{r_0} \theta_0 \frac{\alpha}{r_0 K_0}, \frac{C_4}{\sigma_0^{2(\nu+d)-1}} \right\} \frac{K_0}{\alpha^2} \sigma_0^{2(\nu+d)-1} \leq \varepsilon^2 L_0 (3 \sigma_0^{-1}) \leq 1. (3.33)
\]
Therefore, Lemma 1 implies Lemma 2. \( \square \)

3.2.2. Subsequent steps, iteration and convergence

For \( j \geq 1 \), define
\[
\epsilon_j := \frac{K_0 \epsilon^{2/j} P_j}{\alpha^2}, \quad P_{j+1} := \lambda_s \theta^{j-1} \frac{K_0 P_j}{\alpha^2}, \quad \tilde{\epsilon}_j := \lambda_s \theta^{j} \epsilon_j.
\]
Thus, for any \( j \geq 1 \), one has
\[
\tilde{\epsilon}_{j+1} = \lambda_s \theta^{j+1} \epsilon_{j+1} = \lambda_s \theta^{j+1} \frac{K_0 \epsilon^{2/j+1} P_{j+1}}{\alpha^2} = \lambda_s \theta^{j+1} \frac{K_0 \epsilon^{2/j+1} P_{j+1}}{\alpha^2} = \lambda_s \theta^{j+1} \frac{K_0 \epsilon^{2/j+1} P_j}{\alpha^2} \lambda_s \theta^{j-1} \frac{K_0 P_j}{\alpha^2},
\]
i.e.,
\[
\tilde{\epsilon}_j = \epsilon_j^{2/j-1}.
\]
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.

**V. I. ARNOLD’S “POINTWISE” KAM THEOREM**

Vol. 24 No. 6 2019
Lemma 3. Assume (3.26) ÷ (3.29) and

$$C_8 \theta_0^{-\frac{1}{3}} \hat{e}_1 < 1.$$  \hfill (3.34)

Then one can construct a sequence of symplectic transformations

$$\phi_{j-1}: D_{r_j, s_j}(y_j) \to D_{r_{j-1}, s_{j-1}}(y_{j-1}), \quad j \geq 2$$  \hfill (3.35)

so that

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

converges uniformly.

More precisely, $\varepsilon^{2^{j-1}} P_{j-1}$, $\phi^{-1}$ := $\phi_1 \circ \phi_2 \circ \cdots \circ \phi_{j-1}$, $K_{j-1}$, $y_{j-1}$ converge uniformly on $\{y_0\} \times T^d$ to, respectively, $0$, $\phi^*$, $K_*$, $y_*$ and $H_j \circ \phi^* = K_*$ with $\phi^*$ real-analytic for $x \in T^d_*$ and $\det \partial^2 K_*(y_*) \neq 0$.

Finally, the following estimates hold for any $i \geq 1$:

$$\varepsilon^{2i} \|P_i\|_{r_i, s_i, y_i} \leq \varepsilon^{2i} P_i,$$  \hfill (3.37)

$$|y_{i+1} - y_i| \leq \frac{8\sqrt{2} \varepsilon^{2i} P_i}{r_i},$$  \hfill (3.38)

$$|W_i(\phi^* - \text{id})| \leq \frac{2\varepsilon^{d} \hat{e}_1}{3d^2 \theta_*} \quad \text{on} \quad \{y_*\} \times T^d_*. \hfill (3.39)$$

Remark 1. Notice that $P_1$ is actually independent of $\varepsilon$ (and, in particular, of $\log \varepsilon^{-1}$), while $P_j$ for $j \geq 2$ does depend on $\log \varepsilon^{-1}$ through $\lambda_*$. 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 $i \geq 1$,

$$r_{i+1} = \min \left\{ \frac{\alpha}{4d\sqrt{2K_0k^{\nu}}}, \frac{5}{96d\theta_0}, \frac{5}{96d\theta_0} \right\} = \min \left\{ \frac{r_1}{4^{\nu}}, \frac{r_1}{96d\theta_0}, \frac{r_1}{96d\theta_0} \right\} \hfill (3.35)$$

$$= \min \left\{ \frac{r_1}{4^{\nu}}, \frac{5}{96d\theta_0}, \frac{r_1}{96d\theta_0} \right\} \hfill (3.35)$$

$$= \min \left\{ \frac{r_1}{4^{\nu}}, \frac{5}{96d\theta_0}, \frac{5}{96d\theta_0} \right\} \hfill (3.35)$$

$$= \min \left\{ \frac{r_1}{4^{\nu}}, \frac{5}{96d\theta_0}, \frac{5}{96d\theta_0} \right\} \hfill (3.35)$$

$$= \min \left\{ 2^{2\nu}, \frac{96d\theta_0}{5} \right\} \leq \min \left\{ 2^{2\nu}, \frac{96d\theta_0}{5} \right\} \cdot \theta_0 = C_0 \theta_0. \hfill (3.40)$$

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

there exist $j - 1$ symplectic transformations$^{13}$

$$\phi_i: D_{r_{i+1}, s_{i+1}}(y_{i+1}) \to D_{2r_i, s_i}(y_i), \quad \text{for} \quad 1 \leq i \leq j - 1,$$  \hfill (3.41)

$^{13}$Compare (3.4).
and \( j - 1 \) Hamiltonians \( H_{i+1} = H_i \circ \phi_i = K_{i+1} + \varepsilon^{2i+1} P_{i+1} \) real-analytic on \( D_{r_{i+1},s_{i+1}}(y_{i+1}) \) such that, for any \( 1 \leq i \leq j - 1 \),

\[
\begin{align*}
\| \partial_y^2 K_i \|_{r_i,y_i} & \leq K_i, \quad \| T_i \| \leq T_i, \quad \partial_y K_i(y_i) = \omega, \quad \partial_y^2 K_i(y_i) \neq 0, \\
\| P_i \|_{r_i,s_i,y_i} & \leq P_i, \quad \kappa_i \geq 4\varepsilon^{-1} \log \left( \frac{\sigma_i^{2\nu+d-1}}{\varepsilon_i^{3}} \right), \quad \varepsilon^2 L_i \leq \frac{\sigma_i}{3}
\end{align*}
\]

(3.42)

and

\[
\begin{align*}
\partial_y K_{i+1}(y_{i+1}) & = \omega, \quad \partial_y^2 K_{i+1}(y_{i+1}) \neq 0, \quad |y_{i+1} - y_i| \leq \frac{8\sqrt{\varepsilon P_i}}{r_i}, \\
\| T_{i+1} \| & \leq \| T_i \| + T_i \varepsilon^{2i} L_i, \quad \| K_{i+1} \|_{r_{i+1},y_{i+1}} \leq \| K_i \|_{r_i,y_i} + \varepsilon^{2i} P_i, \\
\| \partial_y^2 K_{i+1} \|_{r_{i+1},y_{i+1}} & \leq \| \partial_y^2 K_i \|_{r_i,y_i} + K_i \varepsilon^{2i} L_i, \\
\max \{ \| W_i(\phi_i - \text{id}) \|_{r_i,s_i,y_i} & , \| \partial_x \pi_2(\phi_i - \text{id}) \|_{s_{i+1}} \} \leq d^{-2} \sigma_i^{d-1} \varepsilon^{2i} L_i, \\
\| P_{i+1} \|_{r_{i+1},s_{i+1},y_{i+1}} & \leq P_i L_i.
\end{align*}
\]

Assume \((\mathcal{P}^j)\) for some \( j \geq 2 \) and let us check \((\mathcal{P}^{j+1})\). Fix \( 1 \leq i \leq j - 1 \). Then

\[
\| \partial_y^2 K_i \|_{r_i,y_i} \leq \| \partial_y^2 K_i \|_{r_i,y_i} + K_i \varepsilon^{2i} L_i \leq K_i + K_i \frac{\sigma_i}{3} = K_{i+1} < K_0 \sqrt{2}
\]

and, similarly,

\[
\| T_{i+1} \| \leq T_{i+1},
\]

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

\[
\frac{\alpha}{r_i K_i} > \frac{\alpha}{r_1 K_0 \sqrt{2}} = 4d\kappa_0 ^{\frac{\nu}{2}} > 1,
\]

(3.44)

so that

\[
\varepsilon^{2i} L_i (3\sigma_i^{-1}) = 3 \varepsilon^{2i} P_i \max \left\{ \frac{8d \sqrt{2} T_0 \theta_0 - (\nu + d)}{r_i^2} \sigma_i, C_4 \max \left\{ 1, \frac{\alpha}{r_i K_i} \right\} \frac{K_0}{\alpha^2 \sigma_i^{-2(\nu+d)}} \right\} \sigma_i^{-1}
\]

(3.44)

\[
\leq 3 \varepsilon^{2i} P_i \max \left\{ \frac{8d \sqrt{2} T_0 \theta_0}{r_i^2}, C_4 \frac{1}{\alpha r_i} \right\} \sigma_i^{-2(\nu+d)-1}
\]

\[
= 3 \max \left\{ 8d \sqrt{2} T_0 \theta_0 \frac{\alpha}{r_i}, C_4 \right\} \sigma_i^{-2(\nu+d)-1} \varepsilon^{2i} P_i \frac{1}{\alpha r_i}
\]

(3.31)

\[
\leq 12d \sqrt{2} \max \left\{ 64d^2, C_4 \right\} \sigma_i^{-2(\nu+d)-1} \varepsilon^{2i} P_i \left( \frac{K_0 \varepsilon^{2i} P_i \theta_0^2}{\alpha^2} \right) C_0^{2(\nu+1)} \kappa_0^{\nu}
\]

(3.40)

\[
\leq 12d \sqrt{2} \max \left\{ 64d^2, C_4 \right\} \sigma_i^{-2(\nu+d)-1} \varepsilon^{2i} P_i \left( \frac{K_0 \varepsilon^{2i} P_i \theta_0^2}{\alpha^2} \right) C_0^{2(\nu+1)} \kappa_0^{\nu}
\]

(3.31)

\[
\leq 3d \cdot 2^{\nu+2d+3} \sqrt{2} \max \left\{ 64d^2, C_4 \right\} \sigma_0^{-2(\nu+d)} \left( \frac{2^{\nu+2d+1} C_0^2 \theta_0^2}{\alpha^2} \right)^{i-1}
\]

\[
\times \frac{K_0 \varepsilon^{2i} P_i}{\alpha^2} \left( \log \frac{\varepsilon_i}{\lambda_i} \right)^{2\nu} \theta_0^2
\]

(3.34)

\[
\leq C_7 \sigma_0^{-2(\nu+2d+1)} \left( \log \frac{\varepsilon_i}{\lambda_i} \right)^{2\nu} \theta_0^2 \theta_*^{i-1} = \lambda_* \theta_*^{i-1} \frac{\varepsilon_i}{\theta_*} = \frac{\hat{\varepsilon}_i}{\theta_*} < 1.
\]
Moreover,
\[ \varepsilon^{2i} L_i < \lambda_i \theta_i^{i-1} \epsilon_i . \]
Thus, by the last relation in (3.43), for any \( 1 \leq i \leq j - 1 \),
\[ \varepsilon^{2i+1} \| P_{i+1} \| \| r_{i+1}, s_i, y_{i+1} \| \leq \varepsilon^{2i} L_i \varepsilon^{2i} P_i < \lambda_i \theta_i^{i-1} \epsilon_i \varepsilon^{2i} P_i = \varepsilon^{2i+1} P_{i+1} , \]
which proves the fourth relation in (3.42) for \( i = j \). Furthermore, by exactly the same computation as above, one gets
\[ \varepsilon^{2i+1} L_{i+1} (3 \sigma_i^{-1}) \leq \frac{\hat{\epsilon}_{i+1}}{\theta_i} = \frac{\varepsilon^{2i}}{\theta_i} < 1 , \]
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 \( H_i, 1 \leq i \leq j \) and get (3.43) and, consequently, (\( S^{2j+1} \)). In fact, we have\(^{14}\)
\[ \lambda_i \theta_i \epsilon_i^2 < \lambda_i \theta_i \epsilon_0 = \epsilon_1 \leq C_7 \sigma_0^{-(4 \nu + 2d + 1)} \theta_i \theta_0^2 \epsilon_0 , \]  
so that
\[ 4 \sigma_j^{-1} \log \left( \sigma_j^{2
u + d} \epsilon_j^{-1} \right) \leq 4 \sigma_j^{-1} \log \left( \epsilon_j^{-1} \right) = 4 \sigma_j^{-1} \log \left( \lambda_i \theta_i \epsilon_1 \epsilon_1^{2j-1} \right) \]
\[ \leq 4 \sigma_j^{-1} \log \left( \lambda_i \theta_i \epsilon_0 \right) \leq 4 \sigma_j^{-1} \log \left( \epsilon_0^{-2j} \right) \]
\[ = 4^j \cdot 4 \sigma_0^{-1} \log (\epsilon_0^{-1}) = \kappa_j . \]
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 \( i \geq 1 \), one needs only to check (\( S^2 \)). 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 (\( S^2 \)).
Next, we prove that \( \phi^j \) is convergent by proving that it is a Cauchy sequence. For any \( j \geq 4 \), we have, using again Cauchy’s estimate (and noting that \( 2^{i-1} \geq i, \forall i \geq 0 \)),
\[ \| W_{j-1}(\phi^{j-1} - \phi^{j-2}) \|_{r_j, s_j, y_j} = \| W_{j-1} \phi^{j-2} \circ \phi_j - W_{j-1} \phi^{j-2} \|_{r_j, s_j, y_j} \]
\[ \leq \| W_{j-1} D \phi^{j-2} W_{j-1}^{-1} \|_{2r_j, s_j, y_j} \| W_{j-1}(\phi_j - \text{id}) \|_{r_j, s_j, y_j} \]
\[ \leq \max \left( r_j, 3 \right) \| W_{j-1} \phi^{j-2} \|_{r_j, s_j, y_j} \times \| W_{j-1}(\phi_j - \text{id}) \|_{r_j, s_j, y_j} \]
\[ = \frac{3}{2 \sigma_{j-1}} \| W_{j-1} \phi^{j-2} \|_{r_j, s_j, y_j} \| W_{j-1}(\phi_j - \text{id}) \|_{r_j, s_j, y_j} \]
\[ \leq \frac{1}{2} \| W_{j-1} \phi^{j-2} \|_{r_j, s_j, y_j} \cdot \sigma_{j-1}^{d} \left( 2^{2j-1} L_j \sigma_{i-1} \right) \]
\[ \leq \frac{1}{2} \| W_{j-1} \phi_1 \|_{r_j, s_j, y_j} \cdot \sigma_{j-1}^{d} \cdot \epsilon_{j-1} \]
\[ \leq \frac{1}{2} \left( \prod_{i=1}^{j-2} \| W_{i+1}^{-1} \| \right) \| W_{1} \phi_1 \|_{r_j, s_j, y_j} \cdot \sigma_{j-1}^{d} \cdot \epsilon_{j-1} . \]

\(^{14}\) Notice that \((\log t)^{2s} \leq t^{1/2}, \forall t \geq e, \forall s \geq 1/4\), so that \( \epsilon_0 (\log \epsilon_0^{-1})^{2\nu} \leq \sqrt{\epsilon_0} \leq e^{-1/2} < 1\), which in turn proves the r.h.s. inequality in (3.45).

\(^{15}\) Observe that for \( j = 2, i = 1 \).
\[
\begin{aligned}
&\frac{1}{2} \left( \prod_{i=1}^{j-2} \frac{r_i}{r_{i+1}} \right) \| W_1 \phi_1 \| r_{2,s_2,y_2} \cdot \sigma_{j-1}^{d} \hat{\epsilon}_{j-1} \\
&= \frac{r_1}{2r_{j-1}} \| W_1 \phi_1 \| r_{2,s_2,y_2} \cdot \sigma_{j-1}^{d} \hat{\epsilon}_{j-1} \\
&\leq \frac{1}{2} \sigma_3^d (C_6 \theta_0)^2 \| W_1 \phi_1 \| r_{2,s_2,y_2} \cdot \left( 2^{-d} C_6 \theta_0 \right)^{j-4} \cdot \hat{\epsilon}_1^{2j-2} \\
&\leq \frac{1}{2} \sigma_3^d (C_6 \theta_0)^2 \| W_1 \phi_1 \| r_{2,s_2,y_2} \cdot \left( 2^{-d} C_6 \theta_0 \right)^{2j-5} \cdot \hat{\epsilon}_1^{2j-2} \\
&= \frac{1}{2} \sigma_3^d (C_6 \theta_0)^2 \| W_1 \phi_1 \| r_{2,s_2,y_2} \cdot \left( 2^{-d} C_6 \theta_0 \right) \frac{1}{2} \hat{\epsilon}_1 \\
&= \frac{1}{2} \sigma_3^d (C_6 \theta_0)^2 \| W_1 \phi_1 \| r_{2,s_2,y_2} \cdot \left( C_8 \theta_0^\frac{1}{2} \hat{\epsilon}_1 \right)^{2j-2}.
\end{aligned}
\]

Therefore, for any \( n \geq 1, j \geq q_0, \)

\[
\| W_1 (\phi^{n+j+1} - \phi^n) \| r_{n+j+2,s_{n+j+2},y_{n+j+2}} \leq \sum_{i=n}^{n+j} \| W_1 (\phi^{i+1} - \phi^i) \| r_{i+2,s_{i+2},y_{i+2}}
\]

\[
\leq \sum_{i=n}^{n+j} \left( \prod_{k=1}^{i} \| W_k W_{k+1}^{-1} \| \right) \| W_{i+1} (\phi^{i+1} - \phi^i) \| r_{i+2,s_{i+2},y_{i+2}}
\]

\[
\leq \sum_{i=n}^{n+j} \left( \prod_{k=1}^{i} \| W_k W_{k+1}^{-1} \| \right) \| W_{i+1} (\phi^{i+1} - \phi^i) \| r_{i+2,s_{i+2},y_{i+2}}
\]

\[
= \sum_{i=n}^{n+j} \| W_{i+1} (\phi^{i+1} - \phi^i) \| r_{i+2,s_{i+2},y_{i+2}}
\]

\[
\leq \frac{1}{2} \sigma_3^d (C_6 \theta_0)^2 \| W_1 \phi_1 \| r_{2,s_2,y_2} \cdot \sqrt{\epsilon} \sum_{i=n}^{n+j} \left( C_8 \theta_0^\frac{1}{2} \hat{\epsilon}_1 \right)^{2i}.
\]

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

To estimate \( |W_0 (\phi^* - \text{id})| \) on \( \{ y_* \} \times T^d_{s_*} \), observe that, for \( i \geq 1, \)

\[
\sigma_i^d \epsilon^{2i} L_i \leq \frac{\sigma_0^{d+1}}{3 \cdot 2^{(d+1)}} \hat{\epsilon}_1^{2i-1} \leq \frac{\sigma_0^{d+1}}{3 \cdot 2^{(d+1)} \theta_*} \hat{\epsilon}_1^i = \left( \frac{2 \sigma_0^{d+1}}{3 \theta_*} \right) \left( \frac{1}{2^d} \hat{\epsilon}_1 \right)^{i+1}
\]

and therefore

\[
\sum_{i \geq 1} \sigma_i^d \epsilon^{2i} L_i \leq \frac{(2 \sigma_0)^{d+1}}{3 \theta_*} \sum_{i \geq 1} \left( \frac{\hat{\epsilon}_1}{2^{d+1}} \right)^i \leq \frac{2 \sigma_0^{d+1}}{3 \theta_*} \hat{\epsilon}_1.
\]

Moreover, for any \( i \geq 1, \)

\[
\| W_1 (\phi^i - \text{id}) \| r_{i+1,s_{i+1},y_{i+1}} \leq \| W_1 (\phi^i - \phi_i) \| r_{i+1,s_{i+1},y_{i+1}} + \| W_1 (\phi_i - \text{id}) \| r_{i+1,s_{i+1},y_{i+1}}
\]

\[
\leq \| W_1 (\phi^i - \phi_i) \| r_{i+1,s_{i+1},y_{i+1}} \left( \prod_{j=0}^{i-1} \| W_j W_{j+1}^{-1} \| \right) \| W_i (\phi_i - \text{id}) \| r_{i+1,s_{i+1},y_{i+1}}
\]

\[
= \| W_1 (\phi^i - \phi_i) \| r_{i+1,s_{i+1},y_{i+1}} \| W_i (\phi_i - \text{id}) \| r_{i+1,s_{i+1},y_{i+1}}
\]

\[
= \| W_1 (\phi^i - \phi_i) \| r_{i+1,s_{i+1},y_{i+1}} \| W_i (\phi_i - \text{id}) \| r_{i+1,s_{i+1},y_{i+1}}
\]

\[
\leq \| W_1 (\phi^i - \phi_i) \| r_{i+1,s_{i+1},y_{i+1}} \| W_i (\phi_i - \text{id}) \| r_{i+1,s_{i+1},y_{i+1}}
\]

\[
\leq \| W_1 (\phi^i - \phi_i) \| r_{i+1,s_{i+1}+d^{-2} \sigma_t^{d-1}} \epsilon^{2i} L_i.
\]
which iterated yields
\[
\|W_1(\phi^i - \text{id})\|_{r_{i+1}, s_{i+1}, y_{i+1}} \leq d^{-2} \sum_{k \geq 1} \sigma_k^{d-1} \varepsilon_k^2 L_k \leq \frac{2\sigma_0^d \varepsilon_1}{3d^2 \theta^*}.
\]
Therefore, taking the limit over \(i\) completes the proof of (3.39) and hence of Lemma 3. \(\square\)

**CONCLUSION**

We can now complete the proof of Theorem A. Let
\[
\begin{align*}
C_{10} & := \left(2^{-(4\nu+2d+1)} + 2C_7\right) C_9/(3d^2), \quad C_{11} := \frac{1}{2^{5\nu+3d-2}} + \frac{C_7 C_9}{3 \cdot 5 \cdot 2^{\nu+2} \cdot d^2 \cdot \sqrt{2}}, \\
C_{12} & := 2^{2\nu+2d+1} C_7^2 C_8 C_9, \quad C_{13} := C_{10} + 2^{-(\nu+1)} C_{11}, \quad C_{14} := 2^{2(3\nu+2d+1)} C_{12}, \\
C_{15} & := 18d^3 + 70, \quad C := 2^{6\nu+3d+8} C_{13}, \quad C_* := \max\left\{(4\nu e^{-1})^{8\nu/3} C_{14}^{2/3}, 2C_{15}C\right\}.
\end{align*}
\]
Observe that
\[
(\log t)^{4\nu} \leq (4\nu e^{-1})^{4\nu} \sqrt{t}, \quad \forall \ t > 1.
\] (3.46)
Then
\[
C_8 \theta_0^\frac{1}{3} \varepsilon_1 = C_{14} \theta^{41/8} (s - s^*)^{-2(3\nu+2d+1)} \varepsilon^2 (\log e^{-1})^{2\nu}
\]
\[
\leq (4\nu e^{-1})^{4\nu} C_{14} \theta^{41/8} \varepsilon^{3/2} (s - s^*)^{-2(3\nu+2d+1)}
\]
\[
< \left(C_* \theta^4 (s - s^*)^{-(6\nu+3d+2)}\right) \varepsilon^{3/2}
\]
(2.4)
\[
\leq 1
\]
and
\[
\varepsilon_0 < C_* \theta^4 (s - s^*)^{-(6\nu+3d+2)} \varepsilon \leq 1.
\]
Hence, (2.4) implies the smallness conditions (3.24) and (3.34). Therefore, Lemmas 2 and 3 hold. Now set \(\phi_* := \phi \circ \phi^*\) and observe that, uniformly on \(\{y_*\} \times \mathbb{T}^d\),
\[
|W_0(\phi_* - \text{id})| \leq |W_0(\phi \circ \phi^* - \phi^*)| + |W_0(\phi^* - \text{id})|
\]
\[
\leq ||W_0(\phi_0 - \text{id})||_{r_1, s_1, y_1} + ||W_0 W_1^{-1}|| \ |W_1(\phi^* - \text{id})|
\]
\[
\leq \frac{1}{d^2} \sigma_0^d \varepsilon L_0 + \frac{2\sigma_0^d}{3d^2 \theta^*} \varepsilon_1 \leq \frac{2\sigma_0^d}{3d^2 \theta^*} C_7 \sigma_0^{-(4\nu+2d+1)} \theta^2 \varepsilon_0
\]
\[
\leq \left(\frac{2C_7}{3d^2}\right) \sigma_0^{-(4\nu+2d+1)} \theta^2 \varepsilon_0 = C_{10} \sigma_0^{-(6\nu+3d+2)} \theta^2 \varepsilon_0 =: \gamma.
\]
Moreover, for any \(i \geq 1\),
\[
|y_i - y_0| \leq \sum_{j=0}^{i-1} |y_{j+1} - y_j| \leq \frac{8T_0 \varepsilon P_0}{r_0} + \sum_{j=1}^{i-1} \frac{8\sqrt{2T_0} \varepsilon^{2^j} P_j}{r_j}
\]
\[
\leq \frac{8T_0 \varepsilon P_0}{r_0} + \sum_{j=1}^{i} \frac{r_j}{10d \theta_0} \sigma_j^{\nu + 2^j \varepsilon} L_j \leq \frac{8T_0 \varepsilon P_0}{r_0} + \frac{r_1}{10d \theta_0} \sigma_1^{\nu} \sum_{j=1}^{i} \sigma_j^{\nu + 2^j \varepsilon} L_j
\]
\[
\leq \frac{8T_0 \varepsilon P_0}{r_0} + \frac{r_1}{10d \theta_0} \sigma_1^{\nu + 2^j \varepsilon} L_j \leq C_{11} \sigma_0^{-(5\nu+3d+1)} \theta^2 \varepsilon P_0 \alpha,
\]
and then, passing to the limit, we get
\[
|y_* - y_0| \leq C_{11} \sigma_0^{-(5\nu+3d+1)} \theta^2 \varepsilon P_0 \alpha.
\]
Thus, the triangle inequality gives
\[ \sup_{T^{n}_u} |W_0(\phi_s - \phi_e)| \leq C_{13} \sigma_0^{-(6\nu + 3d + 2)} \theta^3 \epsilon \],
which proves the bounds on \( u_s \) and \( v_s \) in (2.6). Let us now prove the bound on \( \partial_x u_s \) in (2.6). Set
\[ \tilde{u}_j := \partial_x \pi_2(\phi_j - id), \quad U^j := \partial_x \pi_2 \phi_0 \circ \phi^j = (\mathbb{1}_d + \tilde{u}_0) \circ \cdots \circ (\mathbb{1}_d + \tilde{u}_j). \]
Then, for any \( j \geq 0 \), we have
\[ \|U^j\|_{s+1} \leq (1 + \|\tilde{u}_0\|_{s_0}) \cdots (1 + \|\tilde{u}_j\|_{s_j}) \overset{(3.30)+(3.43)}{\leq} \exp \left( d^{-2} \sum_{k \geq 0} \sigma_k^{d-1} e^{\epsilon k} L_k \right) \leq e^{\gamma}, \]
so that
\[ \|U^{j+1} - U^j\|_{s+1} = \|U^j (\mathbb{1}_d + \tilde{u}_{j+1}) - U^j\|_{s+1} \leq \|U^j\|_{s+1} \|\tilde{u}_{j+1}\|_{s_{j+1}} \overset{(3.30)+(3.43)}{\leq} e^{\gamma} d^{-2} \sigma_{j+1}^{d-1} e^{\epsilon j} L_{j+1}, \]
which implies
\[ \|U^j - \mathbb{1}_d\|_{s+1} \leq e^{\gamma} d^{-2} \sum_{k \geq 0} \sigma_k^{d-1} e^{\epsilon k} L_k \leq \gamma e^{\gamma} L \overset{(2.4)}{\leq} \frac{1}{2} \]
and then, letting \( j \to \infty \), we get the estimate on \( \partial_x u_s \).

**Remark 2.** As it is easy to check, Theorem A holds under the milder condition \( \epsilon \leq \epsilon_\sharp \) where
\[ \epsilon_\sharp := \max \left\{ \epsilon > 0 : C_{14} \theta^{\frac{d}{s}} (s - s_s)^{-2(3\nu + 2d + 1)} \epsilon^2 (\log \epsilon)^{-2\nu} \leq 1, \quad \text{and} \quad 2C (s - s_s)^{-(6\nu + 3d + 2)} \theta^3 \epsilon \exp \left( C (s - s_s)^{-(6\nu + 3d + 2)} \theta^3 \epsilon \right) \leq 1 \right\}. \]
Notice that \( \epsilon_s < \epsilon_\sharp \).
Indeed, the condition
\[ C_{14} \theta^{\frac{d}{s}} (s - s_s)^{-2(3\nu + 2d + 1)} \epsilon^2 (\log \epsilon)^{-2\nu} \leq 1 \]
guaranties the convergence of Arnold’s scheme, while the condition
\[ 2C (s - s_s)^{-(6\nu + 3d + 2)} \theta^3 \epsilon \exp \left( C (s - s_s)^{-(6\nu + 3d + 2)} \theta^3 \epsilon \right) \leq 1 \]
ensures that the torus \( T_{\omega, \epsilon} \) 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 \( \nu \geq d - 1 \geq 1 \) and notice that all \( C_i \)'s are greater than 1 and depend only upon \( d \) and \( \tau \).
\[ \nu \equiv \tau + 1, \quad C_0 := 4\sqrt{2} \left( \frac{3}{2} \right)^{2d+2} \int_{\mathbb{R}^d} \left| \frac{y}{|y|} \right| e^{-|y|} dy, \quad C_1 := 2 \left( \frac{3}{2} \right)^{2d+2} \int_{\mathbb{R}^d} \left| \frac{y}{|y|} \right| e^{-|y|} dy, \]
\[ C_2 := 2^{3d}, \quad C_3 := (2^{d} C^2 + 6d C_1 + C_2) \sqrt{2}, \quad C_4 := \max \left\{ 6d^2 C_0, C_3 \right\}, \]
\[ C_5 := \frac{3 \cdot 2^5 d}{5}, \quad C_6 := \max \left\{ 2^{2d}, C_5 \right\}, \quad C_7 := 3d \cdot 2^{6\nu + 2d + 3} \sqrt{2} \max \left\{ 640d^2, C_4 \right\}, \]
\[ C_8 := \left( 2^{d} C_6 \right)^{\frac{1}{2}}, \quad C_9 := 3 \max \left\{ 80d \sqrt{2}, C_4 \right\}, \quad C_{10} := \left( 2^{-4(4\nu + 2d + 1)} + 2C_7 \right) C_0/(3d^2), \]
\[ C_{11} := \frac{1}{2^{5\nu + 3d + 2}} \frac{C_7 C_9}{3 \cdot 5 \cdot 2^{d + 2} \cdot d \cdot \sqrt{2}}, \quad C_{12} := 2^{2\nu + 2d + 1} C_6^2 C_7 C_8 C_9, \]
\[ C_{13} := C_{10} + 2^{-(\nu + 1)} C_{11}, \quad C_{14} := 2^{2(3\nu + 2d + 1)} C_{12}, \quad C_{15} := 18d^3 + 70, \]
\[ C := 2^{6\nu + 3d + 8} C_{13}, \quad C_s := \max \left\{ (4 \nu e^{-1})^{8\nu/3} C_{14}^{2/3}, 2C_{15} C_s \right\}. \]
APPENDIX B. KOLMOGOROV’S NON-DEGENERACY

Let
\[ \hat{\epsilon} := 2e C (s - s_*)^{-\left(6\tau + 3d + 8\right)} \theta^3 \cdot \frac{K \varepsilon P}{\alpha^2}. \]

Since \( \|\partial_x u_*\|_{s_*} \overset{(2.6)}{\leq} 2/2 \), it follows that \( \text{id} + u_* \) is a diffeomorphism of \( \mathbb{T}^d \). Letting
\[ (\partial_x (\text{id} + u_*) (x))^{-1} =: 1_d + A(x), \]
we have
\[ \|A\|_{s_*} \leq 2 \|\partial_x u_*\|_{s_*} \overset{(2.6)}{\leq} 2 \hat{\epsilon} \leq 1; \quad \|v_*\|_{s_*} \leq 2 \frac{\alpha}{K^2} \frac{\hat{\epsilon}}{\varepsilon P} \leq \frac{r}{4} \theta \leq \frac{r}{8}. \] (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
\[ E = K(y_0), \quad \langle Q_{yy}(0, \cdot) \rangle = K_{yy}(y_0) + \langle M \rangle, \]
\[ M := \partial_y \left( K(y_0 + v_* + y + A^T y) - \frac{1}{2} y^T K_{yy}(y_0)y \right) \bigg|_{y=0} + \partial_y (\varepsilon P \circ \phi) \bigg|_{y=0}, \]
\[ \sup_{y \in \mathbb{T}^d_*} \|K_{yy}(y_0)^{-1} M\| \leq (18d^3 + 70) \hat{\epsilon} \theta \leq 1/2, \]
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 \( |l| + |k| \leq p \) and for any \( 0 < r' < r, 0 < s' < s, \) \(16)\)
\[ \|\partial_y^l \partial_x^k f\|_{r', s'} \leq p! \|f\|_{r,s}(r - r')^{|l|}(s - s')^{|k|}. \]
(ii) For any \( k \in \mathbb{Z}^d \) and any \( y \in D_r(y_0) \)
\[ \|f_k(y)\| \leq e^{-|k|^1} \|f\|_{r,s}. \]

\[16\) As usual, \( \partial_y^l := \frac{\partial^{l_1}}{\partial y_1^{l_1}} \cdots \frac{\partial^{l_d}}{\partial y_d^{l_d}}, \forall y \in \mathbb{R}^d, l \in \mathbb{Z}^d. \]
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\footnote{Here, $D^n_r(z_0)$ denotes the ball in $\mathbb{C}^n$ centred at $z_0$ and with radius $r$.}

\[
F: (y, x) \in D^n_r(y_0) \times D^m_s(x_0) \subset \mathbb{C}^{n+m} \rightarrow F(y, x) \in \mathbb{C}^n
\]

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

\[
\sup_{D^n_r(y_0) \times D^m_s(x_0)} \|1_n - TF(y, x)\| \leq c < 1 \quad \text{and} \quad \sup_{D^n_r(y_0)} |F(y_0, \cdot)| \leq \frac{(1-c)r}{\|T\|}. \quad (C.1)
\]

Then there exists a unique continuous function $g: D^m_s(x_0) \rightarrow D^n_r(y_0)$ 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

\[
\sup_{D^n_r(y_0)} |g - y_0| \leq \frac{\|T\|}{1-c} \sup_{D^m_s(x_0)} |F(y_0, \cdot)|. \quad (C.2)
\]

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 Neishtadt, 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., and Chierchia, L., 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. 1, 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.
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.