J Nonlinear Sci (2017) 27:495–530
DOI 10.1007/s00332-016-9342-5

Journal of Nonlinear Science

A New Approach to the Parameterization Method for Lagrangian Tori of Hamiltonian Systems

Jordi Villanueva[1]

Received: 16 March 2016 / Accepted: 4 October 2016 / Published online: 18 October 2016

Abstract We compute invariant Lagrangian tori of analytic Hamiltonian systems by the parameterization method. Under Kolmogorov's non-degeneracy condition, we look for an invariant torus of the system carrying quasi-periodic motion with fixed frequencies. Our approach consists in replacing the invariance equation of the parameterization of the torus by three conditions which are altogether equivalent to invariance. We construct a quasi-Newton method by solving, approximately, the linearization of the functional equations defined by these three conditions around an approximate solution. Instead of dealing with the invariance error as a single source of error, we consider three different errors that take account of the Lagrangian character of the torus and the preservation of both energy and frequency. The condition of convergence reflects at which level contributes each of these errors to the total error of the parameterization. We do not require the system to be nearly integrable or to be written in action-angle variables. For nearly integrable Hamiltonians, the Lebesgue measure of the holes between invariant tori predicted by this parameterization result is of $\mathcal{O}(\varepsilon^{1/2})$, where ε is the size of the perturbation. This estimate coincides with the one provided by the KAM theorem.

Communicated by Amadeu Delshams.

This work has been partially supported by the Spanish MINECO-FEDER Grant MTM2015-65715-P and the Catalan Grant 2014SGR504.

✉ Jordi Villanueva
jordi.villanueva@upc.edu
http://www.mat.upc.edu/en/people/jordi.villanueva

[1] Departament de Matemàtiques, Universitat Politècnica de Catalunya, Diagonal 647, 08028 Barcelona, Spain

Keywords Hamiltonian systems · KAM theory · Lagrangian tori · Parameterization methods

Mathematics Subject Classification 37J40 · 70K43

1 Introduction

The study of the existence of Lagrangian invariant tori of Hamiltonian systems, carrying quasi-periodic motion, is an outstanding problem of the dynamical systems since the pioneering work of Kolmogorov. In Kolmogorov (1979), Kolmogorov considered a real analytic integrable Hamiltonian for which the frequency map verifies appropriate non-resonance and non-degeneracy conditions for some value of the action variables. If the system has r degrees of freedom, the non-resonance condition is that the corresponding frequency vector $\omega \in \mathbb{R}^r$ is Diophantine, i.e.

$$|\langle k, \omega \rangle| \geq \gamma |k|_1^{-\nu}, \qquad \forall k \in \mathbb{Z}^r \setminus \{0\}, \tag{1}$$

for some $\nu \geq r-1$ and $\gamma > 0$, where $|k|_1 = |k_1| + \cdots + |k_r|$. Hence, the integrable Hamiltonian has a Lagrangian invariant torus carrying quasi-periodic motion with frequencies ω. Non-degeneracy means that the frequency map of the integrable system is a local diffeomorphism between actions and frequencies around the selected action. This latter fact is usually referred to as Kolmogorov's non-degeneracy condition. Kolmogorov's theorem states that this torus persists, slightly deformed, under small analytic Hamiltonian perturbations. The invariant torus of the nearly integrable system also carries quasi-periodic motion with frequencies ω. The proof of this result is performed in Kolmogorov (1979) through the application of a sequence of canonical transformations to the nearly integrable system, with (almost) quadratic rate of convergence. This process leads to a limit Hamiltonian with a specific form—Kolmogorov's normal form—in which the invariant torus also corresponds to a constant value of the action variables.

Assuming that the frequency map of the integrable system defines a diffeomorphism in an open set for the action variables, Arnol'd addressed in Arnold (1963) a global version of Kolmogorov's theorem. Since Moser investigated in Moser (1962) the persistence of quasi-periodic invariant curves of finitely differentiable planar maps, the result of Arnold (1963) is referred to as the KAM theorem and is regarded as one of the most celebrated results of the mechanics. The extension of the KAM theorem to several contexts, as well as the investigation of related results, has led to a very fruitful field of research. There are several nice surveys of KAM theory in the literature (e.g. Broer et al. 2010; de la Llave 2001).

The KAM theorem states that there is a Cantor-like subset of the phase space filled by Lagrangian invariant tori of the nearly integrable Hamiltonian. These tori carry quasi-periodic motion and frequency change with the torus. The Lebesgue measure of the set defined by the holes between invariant tori is bounded from above by $\mathcal{O}(\varepsilon^{1/2})$, where ε is the size of the non-integrable perturbation (e.g. Neĭshtadt 1981; Pöschel 1982). This estimate cannot be improved in general. The proof of the KAM theorem in Arnold (1963) is also performed through a quadratically convergent sequence of canonical transformations. The limit Hamiltonian of this process is integrable, but it is only defined for a Cantor set of actions. The frequencies associated with the limit Hamiltonian verify the Diophantine conditions (1), with γ bounded from below by $\mathcal{O}(\varepsilon^{1/2})$. Therefore, within the framework of the KAM theorem, the lower bound on γ guaranteeing persistence of the torus of frequencies ω of the integrable system is directly related to the upper bound of the Lebesgue measure of the complementary of the Cantor set filled by the surviving tori.

The strategy to build the sequence of canonical transformations leading to Kolmogorov's normal form (for a specific torus of the nearly integrable system) is different from the KAM theorem. Kolmogorov's normal form is defined only on a neighbourhood of the selected torus, but it is still given by an analytic Hamiltonian. Hence, it is not uncommon that KAM and Kolmogorov's theorems end up giving lower bounds for γ of different order with respect to ε, when discussing the persistence of a specific torus of the integrable system. Indeed, the reduction of the nearly integrable system to Kolmogorov's normal form requires solving two coupled small divisor equations (usually referred to as cohomological equations) at each iteration. The reduction to an integrable system on a Cantor set of the KAM theorem involves a single small divisor equation at each step. The quantitative estimates of the solution of each small divisor equation involve a division by γ. For this reason, if we study the persistence of a torus of frequencies ω of the integrable system, by the reduction to Kolmogorov's normal form, the natural lower bound for γ ensuring persistence turns out to be of $\mathcal{O}(\varepsilon^{1/4})$. This is the lower bound on γ that follows from Kolmogorov (1979) as well as the one obtained by several authors in subsequent proofs of Kolmogorov's theorem (e.g. Benettin et al. 1984). Indeed, it is necessary to carry out elaborate estimates on the reduction to Kolmogorov's normal form (see Villanueva 2008) to show that it converges under a lower bound for γ of $\mathcal{O}(\varepsilon^{1/2})$.

Since the primary purpose of this work is to adapt to parameterization methods both the result and the methodology of Villanueva (2008), we summarize the main aspects taken into account in Villanueva (2008). First, we note that the distance of a Hamiltonian to Kolmogorov's normal form has two natural sources of error. One error term is zero when the torus defined by a constant value of the actions is invariant. The other error term vanishes when the dynamics on this invariant torus is quasi-periodic. The distance to Kolmogorov's normal is typically controlled by the maximum size of both errors. In Villanueva (2008), these error terms are dealt with separately. Indeed, the proof takes advantage of the triangular structure of the cohomological equations to define a normalized error that behaves appropriately. Second, canonical transformations and transformed Hamiltonians are written using explicit formulas, as compact as possible. Finally, the expressions of the cohomological equations are used to rewrite some components of the errors after each iteration in terms of the errors of the previous step, rather than in terms of the solutions of these equations.

The reduction to Kolmogorov's normal form is better suited to applications than the KAM theorem, for both performing numerical computations and rigorous verification of the persistence of a specific torus (e.g. Gabern et al. 2005; Locatelli and Giorgilli 2001). However, dealing with Kolmogorov's normal form still presents some drawbacks. Kolmogorov's theorem does not require the system to be nearly integrable, since it can be applied to any Hamiltonian defined as a small perturbation of Kolmogorov's normal form. But the Hamiltonian should be expressed in terms of an appropriate system of action-angle coordinates. The celebrated Liouville–Arnol'd theorem ensures that these coordinates exist for any integrable Hamiltonian with bounded solutions. Actually, a set of action-angle coordinates can always be introduced around any Lagrangian torus, so that it corresponds to a constant value of the actions (see Perry and Wiggins 1994). Although this local result is enough within the present context, building a system of action-angle variables around a specific torus may be computationally very expensive. We refer the reader to Jorba and Villanueva (1998) for the explicit construction of a system of action-angle-like coordinates around a periodic orbit of a Hamiltonian system. A second drawback of Kolmogorov's normal form is that we should perform a sequence of canonical transformations. If we are only interested in computing the parameterization of a specific invariant torus, but we do not need the normalized expression of the Hamiltonian around it, introducing action-angle coordinates and using transformation theory appear to be a huge effort in order to obtain the parameterization.

In de la Llave et al. (2005), it introduced a new method for addressing the persistence of Lagrangian tori of analytic Hamiltonian system, that avoids using canonical transformations. This result can be regarded as the extension of ideas previously developed by several authors (e.g. Celletti and Chierchia 1997; Jorba et al. 1999; Moser 1966a, b; Rüssmann 1976; Salamon and Zehnder 1989; Zehnder 1976). The approach of de la Llave et al. (2005) consists in finding the parameterization of the torus by solving the corresponding invariance equation. The parameterization is defined as an embedding τ of the standard torus $\mathbb{T}^r = (\mathbb{R}/2\pi\mathbb{Z})^r$ in the phase space $\mathbb{R}^{2r}$. If $\omega \in \mathbb{R}^r$ is the frequency vector of the torus, then the invariance equation is that the pullback by τ of the dynamics is the quasi-periodic linear flow of $\mathbb{T}^r$ defined by ω. Under Kolmogorov's non-degeneracy condition, this nonlinear PDE for τ is solved by a quasi-Newton method. The proof is constructive and takes advantage of the geometric and dynamical properties of Hamiltonian systems to solve, approximately, the linearized equations of the Newton method. Hence, only the parameterization is corrected iteratively, but not the Hamiltonian. This construction is usually referred to as the parameterization method in KAM theory.

The parameterization method does not require the Hamiltonian to be neither nearly integrable, nor to be written in action-angle variables. Indeed, this approach is referred in de la Llave et al. (2005) as KAM theory without action-angle variables. It is only necessary to know a good enough approximation to the parameterization of the invariant torus. Quantitatively, the convergence condition is satisfied if $\gamma^{-4}\rho^{-4\nu}\varepsilon$ is small enough, i.e. if γ is bounded from below by an expression of $\mathcal{O}(\varepsilon^{1/4})$. Here, ε is the size of the invariance error of the initial approximation and ρ is the width of the strip of analyticity around $\mathbb{T}^r$ of this initial parameterization. If the Hamiltonian is a perturbation of a system (not necessarily integrable) for which an invariant torus is known, then it is easy to realize that ε is directly related to the size of the perturbation.

The presentation of de la Llave et al. (2005) is performed for both Hamiltonian systems and exact symplectic maps (the discrete-time version of Hamilton's mechanics). Later on, the foundation ideas of the parameterization method for Lagrangian tori have been extended to several contexts, such as lower-dimensional (isotropic) tori that are hyperbolic or elliptic, non-twist invariant tori in degenerate systems and dissipative (conformally symplectic) systems (e.g. Calleja et al. 2013; Fontich et al. 2009; González-Enríquez et al. 2014; Luque and Villanueva 2011). The result of de la Llave et al. (2005), combined with analytic smoothing techniques, has been used to address the existence of invariant tori of finitely differentiable symplectic maps (see González and de la Llave 2008). The parameterization method leads to very efficient algorithms for computing invariant tori (e.g. Calleja and de la Llave 2010; Huguet et al. 2012). Moreover, if we want to prove the existence of a specific torus in the framework of computer-assisted proofs (CAPs), parameterization methods perform better than results based on transformation theory (e.g. Haro et al. 2016). In order to carry out the rigorous verification of the result of de la Llave et al. (2005), the Fourier expansion of an approximate parameterization may be obtained by means of any numerical method for computing Lagrangian tori (see Luque and Villanueva 2016 for an overview of methods). For a detailed exposition of the performances of the parameterization method in the KAM theory, we refer the reader to the recent survey (Haro et al. 2016).

In this paper, we present a new approach to the parameterization method for Lagrangian tori of Hamiltonian systems. Our starting point is the same as in de la Llave et al. (2005), but instead of dealing directly with the invariance equation, we address the persistence of the torus in terms of a set of three different equations for the parameterization. Separately, each of these equations is apparent when the torus is invariant. Altogether, these three equations are equivalent to invariance. The iterative scheme is constructed by solving these functional equations simultaneously by a quasi-Newton method. This indirect approach is motivated by our feeling that, if we address the computation of the torus in terms of the invariance equation for of the parameterization, then it is very tough to obtain sharp estimates. Indeed, there are some components of the invariance error after each iteration that are difficult to bound accurately, unless you have accurate bounds of some projections of the invariance error at the previous step.

The selected conditions for the parameterization have very clear geometric and dynamical meaning and are easy to relate to Kolmogorov's normal form. The first one is that the torus belongs to some energy level set of the Hamiltonian. The second one implies that the torus is a Lagrangian manifold. The third condition is that the dynamics on the torus is quasi-periodic. If the Hamiltonian is ε-close to one for which an invariant torus is known, then the size of the error associated with each of the selected equations is of $\mathcal{O}(\varepsilon)$. Roughly speaking, the convergence condition in this perturbative scenario is $\gamma^{-2}\rho^{-2\nu-1}\varepsilon$ sufficiently small. However, the convergence condition of Theorem 2.6 is a little more involved, since it reflects the separate contribution of each of these three errors to the total error of the initial approximation. Besides decomposition of the error, for the presented construction we also adapt the idea of Villanueva (2008) of using expressions as compact as possible for the involved objects. In particular, with compactness we mean that to define the sequence of parameterizations of the quasi-Newton method we replace, where possible, addition of functions by composition of functions. Indeed, the cohomological equations to be solved at each iteration are equivalent to those of de la Llave et al. (2005), but the sequence of parameterizations is not the same.

The contents of the paper are organized as follows. In Sect. 2, we present some basic notations and definitions and we state the main result of the paper (Theorem 2.6). Our approach to the parameterization method is introduced in Sect. 3 (see Remark 3.4 for the description of one step). Section 4 is devoted to the proof of Theorem 2.6.

2 Statement of the Main Result

Let $h = h(z)$ be a real analytic Hamiltonian with r degrees of freedom, where the coordinates $z = (x, y) \in \mathbb{R}^{2r}$ are symplectic with respect to the two-form $dx \wedge dy$. Although we can easily address the problem when some of the variables are angular coordinates (see Remark 2.2), we restrict the presentation to Hamiltonians written in Cartesian coordinates. The crucial point that we want to stress here is that we do not need any angular coordinate inherent to the system to parameterize the tori. We do not assume that h has any specific form. In particular, we do not need to know whether h is nearly integrable or not. The Hamilton equations for h are $\dot{z} = J\nabla h(z)$, where ∇ stands for the gradient operator and J is the matrix representation of the symplectic two-form, i.e.

$$J = \begin{pmatrix} 0 & \mathrm{Id}_r \\ -\mathrm{Id}_r & 0 \end{pmatrix},$$

For further uses, we recall that J verifies $J^\top = -J$ and $J^2 = -\mathrm{Id}_{2r} = -\mathrm{Id}$.

Given a Diophantine frequency vector $\omega \in \mathbb{R}^r$, we are concerned with the existence of an invariant torus of h (of dimension r) carrying quasi-periodic motion with frequencies ω. Specifically, we seek a parameterization of this torus defined on $\mathbb{T}^r = (\mathbb{R}/2\pi\mathbb{Z})^r$. Next, we introduce some notations and basic properties to be used throughout the paper (see de la Llave et al. 2005 for more details).

Definition 2.1 Let $\mathcal{T} \subset \mathbb{R}^{2r}$ be a real analytic torus of dimension r and $\tau : \mathbb{T}^r \to \mathbb{R}^{2r}$ be an analytic parameterization of $\mathcal{T}$ (i.e. τ is an embedding between $\mathbb{T}^r$ and $\mathcal{T} = \tau(\mathbb{T}^r)$). We introduce the matrix functions

$$N(\theta) = (D\tau(\theta))^\top D\tau(\theta), \qquad \Omega(\theta) = (D\tau(\theta))^\top J\, D\tau(\theta), \qquad \theta \in \mathbb{T}^r, \tag{2}$$

where $D\tau$ is the Jacobian matrix of τ. Hence, N is the matrix representation of the pullback by τ of the standard scalar product and Ω is the matrix representation of the pullback by τ of the symplectic form. Specifically, $\tau^*(\langle\cdot,\cdot\rangle)(\theta)[u, v] = u^\top N(\theta)v$ and $\tau^*(dx \wedge dy)(\theta)[u, v] = u^\top\Omega(\theta)v$. Since $\mathcal{T}$ is a non-degenerate manifold, we have that $\det(N(\theta)) \neq 0$, $\forall\theta \in \mathbb{T}^r$. We also note that $N^\top = N$ and $\Omega^\top = -\Omega$. Moreover, since the symplectic form is exact, $dx \wedge dy = d(-y\, dx)$, it is not difficult to verify that $\Omega = (D\alpha)^\top - D\alpha$, where $\alpha = -(D\tau_x)^\top \tau_y$ and $\tau = (\tau_x, \tau_y)$. Here, α is the vector representation of the pullback by τ of the one-form $-ydx$, i.e. $\tau^*(-y\, dx)(\theta) = \sum_{j=1}^r \alpha_j(\theta)\, \mathrm{d}\theta_j$. In particular, we have that $\langle\Omega\rangle_\theta = 0$, where $\langle(\cdot)\rangle_\theta = (2\pi)^{-r}\int_{\mathbb{T}^r}(\cdot)\, \mathrm{d}\theta$ is the average of any function defined on $\mathbb{T}^r$. The torus $\mathcal{T}$ is a Lagrangian manifold iff $\Omega = 0$. In this latter case, the columns of the $2r \times r$-

dimensional matrices $JD\tau N^{-1}$ and $D\tau$ form a symplectic basis of $\mathbb{R}^{2r}$, at every point of the torus.

Remark 2.2 If h is written in action-angle coordinates, $(x, y) \in \mathbb{T}^r \times \mathbb{R}^r$, and the torus $\mathcal{T}$ is homotopic to $\mathbb{T}^r \times \{0\}$, then the parameterization of $\mathcal{T}$ should be of the form $\theta \in \mathbb{T}^r \mapsto (\theta + \tau_x(\theta), \tau_y(\theta))$, where $\tau : \mathbb{T}^r \to \mathbb{R}^{2r}$. The adaptation of Theorem 2.6 to this context is straightforward.

Definition 2.3 We assume that the components of $\omega \in \mathbb{R}^r$ are rationally independent and that the torus $\mathcal{T}$ of Definition 2.1 is an invariant torus of h, carrying quasi-periodic motion with frequencies ω. Then, there is a parameterization τ of $\mathcal{T}$ such that the pull-back by τ of the dynamics of h is the quasi-periodic linear flow of $\mathbb{T}^r$ with frequencies ω. This parameterization is unique if we fix the value of τ at any point and verifies

$$L_\omega \tau(\theta) = J\nabla h(\tau(\theta)), \qquad \forall \theta \in \mathbb{T}^r, \tag{3}$$

where $L_\omega(\cdot) = \sum_{j=1}^{r} \omega_j\, \partial_{\theta_j}(\cdot)$ is the Lie derivative with respect to ω. We refer to this nonlinear system of PDEs as the invariance condition of τ. Equation (3) means that the correspondence $t \in \mathbb{R} \mapsto \tau(\omega t + \theta_0)$ defines a quasi-periodic trajectory of h, $\forall \theta_0 \in \mathbb{T}^r$. Equation (3) also implies that $L_\omega \Omega = 0$ [see (17) and (21)]. Since $\langle \Omega \rangle_\theta = 0$ by Definition 2.1, we have that $\Omega = 0$ [see (5)]. Consequently, any invariant torus of h of dimension r, with quasi-periodic dynamics, is automatically a Lagrangian manifold.

The assertion $\Omega = 0$ of Definition 2.3 follows from the analysis of the small divisor equation

$$L_\omega f(\theta) = g(\theta), \qquad \theta \in \mathbb{T}^r, \tag{4}$$

where f and g may be scalar-valued, vector-valued or matrix-valued. If we expand g in Fourier series,

$$g(\theta) = \sum_{k\in\mathbb{Z}^r} \hat{g}(k)\, \mathrm{e}^{\mathrm{i}\langle k,\theta\rangle},$$

then the average of g, $\langle g \rangle_\theta = \hat{g}(0)$, should be zero if we want a solution for f of (4). If $\langle g \rangle_\theta = 0$, then the Fourier coefficients of f verify $\mathrm{i}\langle k, \omega\rangle \hat{f}(k) = \hat{g}(k)$, $\forall k \in \mathbb{Z}^r$. Consequently, if we introduce the notation $\widetilde{f} = f - \langle f \rangle_\theta$, then there is a unique formal solution for $\widetilde{f}$ of (4) and $\hat{f}(0)$ is free to take any value. Explicitly, any formal solution for f takes the form $f(\theta) = \langle f \rangle_\theta + L_\omega^{-1} g(\theta)$, where

$$L_\omega^{-1} g(\theta) = \sum_{k\in\mathbb{Z}^r\setminus\{0\}} \frac{\hat{g}(k)}{\mathrm{i}\langle k, \omega\rangle}\, \mathrm{e}^{\mathrm{i}\langle k,\theta\rangle}. \tag{5}$$

In the particular case $g = 0$, we conclude that f should be constant. Hence, conditions $L_\omega \Omega = 0$ and $\langle \Omega \rangle_\theta = 0$ altogether imply that $\Omega = 0$ in Definition 2.3.

To ensure the convergence of (5) when $g \neq 0$, we should take into account the regularity of g and the effect of the "small divisors" $\langle k, \omega \rangle$. Lemma 2.4 provides quantitative estimates for $L_\omega^{-1} g$ if g is analytic and ω is Diophantine. To state Lemma 2.4, we introduce the following notations:

$$\Delta(\rho) = \{\theta \in \mathbb{C}^r : |\mathrm{Im}(\theta)| < \rho\}, \qquad \|f\|_\rho = \sup_{\theta \in \Delta(\rho)} |f(\theta)|, \qquad \rho > 0, \tag{6}$$

where $|x|$ is the sup-norm of the complex vector x. If $A \in \mathbb{M}_{n\times m}(\mathbb{C})$, we also denote by $|A|$ the compatible matrix norm. Hence, the set $\Delta(\rho)$ is the complex strip of width ρ around $\mathbb{R}^r$ and $\|f\|_\rho$ the associated sup-norm for a complex-valued function f.

Lemma 2.4 (Rüssmann estimates Rüssmann 1975) *Let g be a complex-valued function defined on $\mathbb{R}^r$ and 2π-periodic in all the variables. We suppose that g can be analytically extended to the interior of $\Delta(\rho)$, for some $\rho > 0$, and that g is bounded in the closure of $\Delta(\rho)$. Let $\omega \in \mathbb{R}^r$ be a Diophantine vector which verifies* (1)*, for some $\gamma > 0$ and $\nu \geq r - 1$. If $\langle g \rangle_\theta = 0$, then* Eq. (4) *has a unique solution $f = L_\omega^{-1} g$, which is analytic in $\Delta(\rho)$, is 2π-periodic in all the variables and has zero average on $\mathbb{T}^r$. This solution is given by* (5) *and verifies the following estimates:*

$$\|L_\omega^{-1} g\|_{\rho-\delta} \leq \sigma \frac{\|g\|_\rho}{\gamma \delta^\nu}, \qquad \forall\, 0 < \delta \leq \rho,$$

where $\sigma \geq 1$ can be taken as

$$\sigma(r, \nu, |\omega|_1/|\omega|) = \frac{3\pi}{2^{\nu+1}} 6^{r/2} \sqrt{\nu \Gamma(2\nu)} \left(\frac{|\omega|_1}{|\omega|}\right)^\nu,$$

where $\Gamma(\cdot)$ stands for the gamma function. Moreover, if g is real analytic, then so is f.

Remark 2.5 To use Lemma 2.4 for a vector-valued (resp., matrix-valued) function g, we extend the notation $\|g\|_\rho$ to it by computing the $|\cdot|$-norm of the constant vector (resp., matrix) defined by the $\|\cdot\|_\rho$-norms of the entries of g. Another advantage of this definition is that if the product of two matrix-valued functions g_1 and g_2 is well defined, then $\|g_1 g_2\|_\rho \leq \|g_1\|_\rho \cdot \|g_2\|_\rho$.

Finally, if $F : \mathcal{U} \subset \mathbb{C}^{2r} \to \mathbb{C}$ is an analytic function defined in an open set $\mathcal{U}$ and bounded in the closure of $\mathcal{U}$, then we introduce the norm

$$\|F\|_{\mathcal{U}} = \sup_{z \in \mathcal{U}} |F(z)|. \tag{7}$$

We extend the notation $\|\cdot\|_{\mathcal{U}}$ to vector-valued and matrix-valued functions as done in Remark 2.5 for the norm $\|\cdot\|_\rho$.

We use notations introduced above to state our main result.

Theorem 2.6 *Let $h : U \subset \mathbb{R}^{2r} \to \mathbb{R}$ be a Hamiltonian with r degrees of freedom, defined in an open set U, and $\tau : \mathbb{T}^r \to \mathbb{R}^{2r}$ a parameterization of a torus of dimension r such that $\mathcal{T} = \tau(\mathbb{T}^r) \subset U$. We suppose that $h = h(z)$ can be analytically extended to the open set $\mathcal{U} \subset \mathbb{C}^{2r}$, with $U \subset \mathcal{U}$, and that $\tau = \tau(\theta)$ can be analytically extended to the complex strip $\Delta(\rho)$, for some $\rho > 0$, with $\tau(\Delta(\rho)) \subset \mathcal{U}$. We assume that the following estimates hold, for some constants $\{m_j\}_{j=1}^6$:*

$$\|D^2 h\|_{\mathcal{U}} \leq m_1, \qquad \|D^2(\partial_{z_j} h)\|_{\mathcal{U}} \leq m_2, \quad j = 1, \ldots, 2r, \tag{8}$$

$$\|D\tau\|_\rho < m_3, \qquad \|N^{-1}\|_\rho < m_4, \qquad d_{\tau,\rho,\mathcal{U}} > m_5 > 0, \tag{9}$$

where $D^2(\cdot)$ is the Hessian matrix of the scalar function $(\cdot)$ and we introduce the notation

$$d_{\tau,\rho,\mathcal{U}} = \mathrm{dist}(\tau(\Delta(\rho)), \partial\mathcal{U}), \tag{10}$$

where $\partial\mathcal{U}$ stands for the boundary of $\mathcal{U}$ and $\mathrm{dist}(\cdot,\cdot)$ denotes the distance between two complex sets (using the sup-norm). Moreover, we introduce the r-dimensional symmetric matrix function S,

$$S(\theta) = (N(\theta))^{-1}(D\tau(\theta))^\top \left[J\, D^2 h(\tau(\theta))\, J + D^2 h(\tau(\theta))\right] D\tau(\theta)\,(N(\theta))^{-1}, \tag{11}$$

and we suppose that $\det(\langle S\rangle_\theta) \neq 0$ and that

$$|(\langle S\rangle_\theta)^{-1}| < m_6. \tag{12}$$

From these quantities, we define $\Theta > 0$ as

$$\Theta = \min\left\{m_3 - \|D\tau\|_\rho, m_4 - \|N^{-1}\|_\rho, d_{\tau,\rho,\mathcal{U}} - m_5, m_6 - |(\langle S\rangle_\theta)^{-1}|\right\}. \tag{13}$$

We consider a Diophantine (frequency) vector $\omega \in \mathbb{R}^r$ that verifies conditions (1), for some $0 < \gamma \leq 1$ and $\nu \geq r - 1$, and we denote by $\sigma = \sigma(r, \nu, |\omega|_1/|\omega|)$ the constant provided by Lemma 2.4. From h, τ and ω, we compute the (error) functions e_1, e_2, and e_3, defined as

$$e_1(\theta) = h(\tau(\theta)) - \langle h(\tau(\theta))\rangle_\theta, \quad e_2(\theta) = \Omega(\theta)\omega,$$
$$e_3(\theta) = (N(\theta))^{-1}(D\tau(\theta))^\top J\nabla h(\tau(\theta)) - \omega, \tag{14}$$

and we suppose that, for some constants $0 \leq \mu_j \leq 1$, $j = 1, 2, 3$, we have

$$\|\nabla e_1\|_\rho \leq \mu_1, \qquad \|e_j\|_\rho \leq \mu_j, \quad j = 2, 3. \tag{15}$$

There is a constant $m \geq 1$, that depends only on r, σ and $\{m_j\}_{j=1}^6$, for which the following result holds. Given $0 < \delta < \min\{1, \rho/10\}$, we suppose that the quantities $\{\mu_j\}_{j=1}^3$ are small enough so that:

$$\frac{m}{\gamma \delta^{\nu+1}} \left(\mu_3 + \frac{\mu_1 + \mu_2}{\gamma \delta^\nu} \right) < \min \left\{ \frac{1}{2^{2(\nu+2)}}, \frac{\Theta}{2} \right\}. \tag{16}$$

Then, there is $\tau^* : \mathbb{T}^r \to \mathbb{R}^{2r}$ giving the analytic parameterization of an invariant torus $\mathcal{T}^*$ of h, carrying quasi-periodic motion with frequencies ω. Specifically, we have that $\mathcal{T}^* = \tau^*(\mathbb{T}^r) \subset U$, that τ^* can be analytically extended to the complex strip $\Delta(\rho^*)$, where $\rho^* = \rho - 10\delta$, with $\tau^*(\Delta(\rho^*)) \subset \mathcal{U}$, and verifies

$$\begin{aligned} L_\omega \tau^*(\theta) &= J \nabla h(\tau^*(\theta)), \qquad \forall \theta \in \mathbb{T}^r, \\ \|\tau^* - \tau\|_{\rho^*} &\le \frac{2m}{\gamma \delta^\nu} \left(\mu_3 + \frac{\mu_1 + \mu_2}{\gamma \delta^\nu} \right), \\ \|D\tau^* - D\tau\|_{\rho^*} &\le \frac{2m}{\gamma \delta^{\nu+1}} \left(\mu_3 + \frac{\mu_1 + \mu_2}{\gamma \delta^\nu} \right). \end{aligned}$$

Moreover, if we denote by N^* and S^* the expressions defined by replacing τ by τ^* in the definitions of N and S, then we have that $\det(\langle S^* \rangle_\theta) \neq 0$ *and*

$$\|D\tau^*\|_{\rho^*} \le m_3, \quad \|(N^*)^{-1}\|_{\rho^*} \le m_4, \quad d_{\tau^*, \rho^*, \mathcal{U}} \ge m_5, \quad |(\langle S^* \rangle_\theta)^{-1}| \le m_6.$$

The proof of Theorem 2.6 is constructive and well suited for both numerical purposes and CAPs (see Remark 3.4). In particular, we do not apply any transformation to h. The cornerstone of Theorem 2.6 is that conditions $\nabla e_1 = e_2 = e_3 = 0$ [see (14)] are equivalent to the invariance condition (3) for τ. We refer the reader to Sects. 3.2 and 4 for details. In fact, to ensure invariance, we should also verify that the columns of the matrix functions $D\tau$ and $J D\tau$ form a basis at any point of the torus. This latter assumption is automatic if μ_2 is small enough, as guaranteed by (16).

Remark 2.7 The non-degeneracy condition $\det(\langle S \rangle_\theta) \neq 0$ that only depends on h and τ was introduced in de la Llave et al. (2005). If $\mathcal{T}$ is an invariant torus, then the matrix S defines the infinitesimal approximation of the frequency map around the torus. Explicitly, we can define a $2r \times r$-dimensional matrix function $T(\theta)$, whose columns are symplectically conjugate directions to the tangent ones of $\mathcal{T}$ at $\tau(\theta)$, with the following property. If we consider the torus $\tau(\theta) + T(\theta)\,\xi$, for small values of $\xi \in \mathbb{R}^r$, then, at first order, this torus is invariant with frequency vector $\omega - \langle S \rangle_\theta\, \xi$. For more details, see the last paragraph of Sect. 3.1. Furthermore, in the action-angle framework, the condition $\det(\langle S \rangle_\theta) \neq 0$ reads as Kolmogorov's non-degeneracy condition when formulated on an invariant torus of a Hamiltonian written in Kolmogorov's normal form. Explicitly, if we consider the Hamiltonian (24) and compute S for the invariant torus $\mathbb{T}^r \times \{0\}$, parameterized by $\tau(\theta) = (\theta, 0)$, then we obtain that $S = -A$.

Remark 2.8 We do not provide any specific value for the constant m of Theorem 2.6, but since in Sect. 4 we display explicit formulas for the expressions involved in the proof, it is not difficult to generate it. In the way we define m, it does not depend on $\|\nabla h\|_\rho$, δ, ρ, γ and on the size of any component of ω. This fact is mainly due to the strategy followed in the proof. When performing a CAP, since the constants in the statement take specific numerical values, it may be recommendable to modify some steps of the proof in order to improve the numerical value of the condition of convergence. Indeed, it may be a good idea adding to the statement a constant m_7 so that $\|D^2(\partial_{\theta_j}\tau)\|_\rho < m_7$, $\forall j = 1, \ldots, r$, and then controlling the size of $D^2(\partial_{\theta_j}\tau)$ iteratively throughout the proof.

Remark 2.9 If h is a ε-small perturbation of a Hamiltonian for which we know the parameterization τ of an invariant torus, then the size of e_j is of $\mathcal{O}(\varepsilon)$, $j = 1, 2, 3$ (in fact, we can set $e_2 = 0$). If h is not a natural perturbation of a simpler system, but we know the size of the invariance error e associated with the parameterization τ of a quasi-torus of h [see (17)], then we can relate each e_j to e [see (36)]. However, bounding e_1 and e_2 in terms of e involves the resolution of a small divisor equation, which means an estimate of the form $\mathcal{O}(\|e\|_\rho/\gamma)$. Hence, when applying Theorem 2.6, it seems more useful dealing with e_1, e_2, and e_3 as specific errors, independent of the invariance error e [see Eqs. (51) and (54) for the expression of e in terms of e_1, e_2, and e_3].

Remark 2.10 It may be interesting to combine Theorem 2.6 with analytic smoothing techniques to extend this parameterization result to finitely differentiable Hamiltonians (compare with González and de la Llave 2008). Since the maximum exponent of δ in the condition of convergence (16) is $2\nu + 1$, our guess is that the minimum regularity that we should ask to h is $\mathcal{C}^\ell$, with $\ell > 2\nu + A$, for some $A \geq 2$ (we observe that the definition of μ_1 involves ∇h). The best case scenario is to be able to set $A = 2$, in accordance with Pöschel (1982), Salamon (2004).

Remark 2.11 If h is ε-close to an integrable Hamiltonian, then we may use Theorem 2.6 to prove the classical KAM theorem, without writting h in action-angle variables. Let $\Omega \subset \mathbb{R}^r$ be the set of frequencies attained by the frequency map of the integrable system, and let $\Omega_\gamma \subset \Omega$ be the Cantor-like set defined by the values $\omega \in \Omega$ that verify the Diophantine conditions (1). Using Theorem 2.6, we may try to set a priori a value $\gamma = \mathcal{O}(\sqrt{\varepsilon})$ so that there is an invariant torus of h, with frequencies ω, for each $\omega \in \Omega_\gamma$. Indeed, we introduce the frequencies $\omega \in \Omega_\gamma$ as a parameter, we consider the family of parameterizations $\omega \mapsto \tau_\omega$ provided by the integrable system as initial approximations, and we apply Theorem 2.6 to prove that there is a family of parameterizations $\omega \mapsto \tau^*_\omega$ giving an invariant torus of h with frequencies ω. To control the measure of the portion of the phase space filled by these invariant tori, it is enough to show that the dependence $\omega \mapsto \tau^*_\omega$ is Lipschitz [e.g. see the elegant proof of the classical KAM theorem performed in Pöschel (2001)]. To carry out the explicit control of the Lipschitz constant of $\omega \mapsto \tau^*_\omega$, we may use analogous strategies as those in Jorba and Villanueva (1997a, b) for proving existence of families of lower-dimensional tori of Hamiltonian systems, of dimension $s < r$, using $\omega \in \mathbb{R}^s$ as a parameter.

Remark 2.12 We are convinced that the approach of Theorem 2.6 can easily be adapted for dealing with conditions of non-degeneracy different from the one of Kolmogorov. The most straightforward case appears to be the iso-energetic KAM theorem. It would also be interesting to consider the extension of this result to dynamical systems that are not Hamiltonian. A very natural context is the Lagrangian tori of exact symplectic maps (under Kolmogorov's non-degeneracy condition). For these discrete dynamical systems, the error functions $\{e_j\}_{j=1}^3$ of the parameterization can be defined in terms of the map, the rotation frequency of the torus and the one-form of the symplectic structure. Details are discussed in a work currently in preparation.

3 Formal Presentation of the Methodology

In this section, we introduce our approach to the parameterization method for Lagrangian tori of analytic Hamiltonian systems. In Sect. 3.1, we outline some of the main aspects of the original construction of de la Llave et al. (2005). This brief summary is useful to set some basic notations and properties to be used throughout the paper, as well as to point out some differences between de la Llave et al. (2005) and the presented method. In Sect. 3.2, the set of equations which define our approach to the parameterization method is introduced. The formal aspects of the quasi-Newton method used to solve these equations are developed in Sect. 3.3 (see Remark 3.4 for a summary).

In what follows, we use definitions and notations introduced in Sect. 2. To simplify notations, from now on we omit the explicit dependence on θ of τ, N, and Ω.

3.1 The Parameterization Method for Lagrangian Tori

Let $\omega \in \mathbb{R}^r$ be a Diophantine frequency vector, fixed from now on. Our aim is to compute the parameterization τ of an invariant torus $\mathcal{T}$ of h, carrying quasi-periodic motion with frequencies ω. Indeed, we should set to zero the invariance error of τ, defined as:

$$e = L_\omega \tau - J\nabla h(\tau). \tag{17}$$

If the norm $\|e\|_\rho$ is "small", for some $\rho > 0$, we say that $\mathcal{T}$ is a quasi-torus of h. Assume known the parameterization τ of a quasi-torus of h. The core idea of parameterization methods is to compute a new parameterization $\tau^{(1)} = \tau + \Delta\tau$, where $\Delta\tau$ is obtained by solving (at least approximately) the linearized invariance equation for $\tau^{(1)}$ around τ. If $e^{(1)}$ is the invariance error for $\tau^{(1)}$, then it should be almost quadratic with respect to e (quadratic modulo the effect of the small divisors associated with ω and the performed Cauchy estimates). If the initial error is small enough, we expect convergence of the iteration of this quasi-Newton method to the parameterization τ^* of an invariant torus of h.

Since e is small, we have that the columns of the matrices $JD\tau\, N^{-1}$, and $D\tau$ form an approximate symplectic basis at any point of $\mathcal{T}$ [see Eq. (21) and comments below on the smallness of Ω]. The method of de la Llave et al. (2005) takes advantage of this fact to look for the improved parameterization $\tau^{(1)}$ as follows:

$$\tau^{(1)} = \tau + \Delta\tau, \qquad \Delta\tau = JD\tau N^{-1}\, a + D\tau\, b, \tag{18}$$

where $a(\theta)$ and $b(\theta)$ are r-dimensional (small) vector functions of $\mathbb{T}^r$ to be determined. If we formulate the invariance condition for $\tau^{(1)}$ and we linearize it around τ, then we obtain the following linear system of PDEs for $\Delta\tau$ (Newton method):

$$\mathcal{R}(\Delta\tau) = -e, \tag{19}$$

where $\mathcal{R}(\cdot) = L_\omega(\cdot) - J D^2 h(\tau(\theta))(\cdot)$ is a linear differential operator acting on vector functions of θ of dimension $2r$ and, more in general, on matrix functions of θ with $2r$ rows. The cornerstone of de la Llave et al. (2005) is to show that the action of $\mathcal{R}$ on the columns of the matrices $J D\tau N^{-1}$ and $D\tau$ takes the following form:

$$\mathcal{R}(D\tau) = E_1, \qquad \mathcal{R}(J D\tau N^{-1}) = -D\tau\, S + E_2, \tag{20}$$

where S is the r-dimensional matrix function defined in (11) and E_1 and E_2 are "small" matrices, with size controlled in terms of the size of e. Consequently,

$$\mathcal{R}(\Delta\tau) = J\, D\tau N^{-1}\, L_\omega a - D\tau\, Sa + D\tau\, L_\omega b + E_3,$$

where E_3 is quadratic with respect to e, a, and b. The crucial point to get formulas in (20) is that the size of the matrix Ω is controlled by the size of e through the Lie derivative $L_\omega \Omega$. Specifically:

$$L_\omega \Omega = (L_\omega(D\tau))^\top J\, D\tau + (D\tau)^\top J L_\omega(D\tau) = (De)^\top J\, D\tau + (D\tau)^\top J\, De, \tag{21}$$

where we have used that L_ω commutes with $D = D_\theta$ to obtain

$$L_\omega(D\tau) = D(L_\omega \tau) = D(D\tau\, \omega) = D(J\nabla h(\tau) + e) = J D^2 h(\tau) D\tau + De. \tag{22}$$

Since definition of Ω implies that $\langle \Omega \rangle_\theta = 0$, the size of Ω can be bounded by applying L_ω^{-1} to the rightmost expression of (21) (see Lemma 2.4). If we multiply Eq. (19) by $(D\tau)^\top J$ and $N^{-1}(D\tau)^\top J$ and we remove the quadratic terms, then we obtain the following (cohomological) equations for a and b:

$$L_\omega a = (D\tau)^\top J e, \qquad L_\omega b = Sa - N^{-1}(D\tau)^\top e. \tag{23}$$

We can solve this triangular system of equations provided that $\det(\langle S \rangle_\theta) \neq 0$ and that $\langle (D\tau)^\top J e \rangle_\theta = 0$. This latter condition is automatic from the definition of e in (17) (see details in de la Llave et al. 2005). The solutions for a and b of (23) are unique up to the value of the average of b, that is free to take any value. Since $\langle b \rangle_\theta$ only influences on the origin of the parameterization, we take $\langle b \rangle_\theta = 0$. By constructing $\tau^{(1)}$ in this way, it is not difficult to realize that $e^{(1)}$ is (almost) quadratic with respect to e. In de la Llave et al. (2005), it is shown that the iteration of this procedure converges to τ^* if $\gamma^{-4}\rho^{-4\nu}\|e\|_\rho$ is sufficiently small, where the values of γ and ν are those of the Diophantine conditions in (1) and $\|e\|_\rho$ is the size of the initial error.

To illustrate the main geometric and dynamical aspects of the construction above, firstly we consider the case in which the torus $\mathcal{T}$ is invariant. This means that e is zero and so are Ω, E_1, and E_2. Then, in de la Llave et al. (2005) it is shown that formulas in (20) imply that the variational equations of h around $\mathcal{T}$ can be reduced to constant coefficients (in the aim of quasi-periodic Floquet theory) by a linear quasi-periodic transformation. The reducing transformation and the reduced matrix can be written explicitly in terms of $D\tau$, $JD\tau N^{-1}$, and S. This property is usually referred to as the automatic reducibility of Lagrangian tori. Explicitly, we compute $B = L_{\omega}^{-1}(S - \langle S \rangle_{\theta})$ and we introduce $T = JD\tau N^{-1} + D\tau B$. The $2r \times 2r$ dimensional matrix function M, defined by joining together the columns of the matrices T and $D\tau$, is a symplectic matrix function of $\mathbb{T}^r$. Then, the reducing transformation is $z = M(\omega t)Z$ and the reduced matrix is given, by blocks, by $\left(\begin{smallmatrix} 0 & 0 \\ -\langle S \rangle_{\theta} & 0 \end{smallmatrix}\right)$. If $\mathcal{T}$ is a quasi-torus of h, then formulas in (20) imply quasi-reducibility of $\mathcal{T}$, also automatically. Although automatic quasi-reducibility is not explicitly involved in our approach to the parameterization method for Lagrangian tori (at least not in a visible manner), it is a crucial aspect of parameterization methods in KAM theory.

3.2 A Modified Approach to the Parameterization Method

In this section, we introduce the equations which define our approach to the parameterization method. The selected set of equations is inspired by Kolmogorov's normal form. Let us suppose, just for a moment, that the Hamiltonian h is written in a symplectic system of action-angle variables, i.e. $h = h(\theta, I)$, where $(\theta, I) \in \mathbb{T}^r \times \mathbb{R}^r$. The Hamilton equations for h are

$$\dot{\theta} = \nabla_I h(\theta, I), \qquad \dot{I} = -\nabla_{\theta} h(\theta, I).$$

We say that h is in Kolmogorov's normal form (at $I = 0$) if it can be written as

$$h(\theta, I) = \lambda + \langle \omega, I \rangle + \frac{1}{2}\langle I, A(\theta) I \rangle + F(\theta, I), \tag{24}$$

for some $\lambda \in \mathbb{R}$, $\omega \in \mathbb{R}^r$, a symmetric matrix function $A(\theta)$, and a scalar function $F = \mathcal{O}_3(I)$. Hence, the torus $\mathbb{T}^r \times \{0\}$ is an invariant manifold of (24) with a linear conditionally periodic flow for θ, i.e. $\dot{\theta} = \omega$. If the components of ω are rationally independent, then the dynamics on the torus is quasi-periodic. Kolmogorov's non-degeneracy condition for this torus reads as $\det(\langle A \rangle_{\theta}) \neq 0$. Synthetically, expression (24) is equivalent to the following conditions on h:

$$\nabla_{\theta} h(\theta, 0) = 0, \qquad \nabla_I h(\theta, 0) = \omega.$$

Consequently, if we want to control the distance of any given Hamiltonian h to the reduced form (24), then we have two different sources of error:

$$e_1(\theta) = h(\theta, 0) - \langle h(\theta, 0) \rangle_{\theta}, \qquad e_2(\theta) = \nabla_I h(\theta, 0) - \omega.$$

The condition $e_1 = 0$ means that the torus $\mathbb{T}^r \times \{0\}$ belongs to an energy level set of h. This condition automatically implies that this torus is invariant by h, regardless of the value of e_2. The reason that $e_1 = 0$ means invariance of $\mathbb{T}^r \times \{0\}$ is that, since we are working with action-angle coordinates, then $\mathbb{T}^r \times \{0\}$ is a Lagrangian manifold in a straightforward way (see Proposition 3.1). The vector condition $e_2 = 0$ implies that the dynamics on this invariant torus is quasi-periodic, with frequencies ω.

Classical proofs of Kolmogorov's theorem define the distance of h to the reduced form (24) as $\varepsilon = \max\{\|e_1\|_\rho, \|e_2\|_\rho\}$ and show convergence of the reduction to Kolmogorov's normal form if $\gamma^{-4}\rho^{-s}\varepsilon$ is sufficiently small. The particular value of $s > 0$ varies depending on the author [e.g. $s = 8(r+1)$ in Benettin et al. (1984)]. In Villanueva (2008), the norms of e_1 and e_2 are dealt with separately and the convergence condition can be written as $\gamma^{-1}\rho^{-\nu-5}(\|e_2\|_\rho + \gamma^{-1}\rho^{-\nu}\|\nabla e_1\|_\rho)$ sufficiently small. This expression reveals the specific role played by each initial error, e_1 and e_2, in the convergence of the method. A convergence condition like this latter one is what we pursued when formulating Theorem 2.6 [see (16)]. We note that, in terms of ε defined as above, the convergence condition of Villanueva (2008) reads as $\gamma^{-2}\rho^{-2\nu-6}\varepsilon$ sufficiently small.

We resume the study of the case in which τ is the parameterization of a quasi-torus $\mathcal{T}$ of a Hamiltonian h written in Cartesian coordinates. With Kolmogorov's normal form in mind, the most significant difference with respect to the action-angle context is the control of how far is $\mathcal{T}$ from being a Lagrangian manifold. In some situations, the first approximation may be a Lagrangian torus by the context. However, preserving this property iteratively, without using canonical transformations, seems beyond the scope of parameterization methods. A natural way to measure this distance is through the norm of the matrix function Ω [see (2)]. As shown in de la Llave et al. (2005), the size of Ω can be related to the size of the invariance error e of τ [see (17)] through the Lie derivative $L_\omega\Omega$ [see (21)]. Proceeding in this way, we end up with an estimate of the form $\|\Omega\|_{\rho-2\delta} = \mathcal{O}(\gamma^{-1}\delta^{-\nu-1}\|e\|_\rho)$, which involves a division by γ that we want to avoid. Actually, using e to control Ω, we obtain the impression that establishing convergence of the parameterization method when γ is bounded from below by $\mathcal{O}(\varepsilon^{1/2})$ appears to be a very difficult task. Hence, one of the crucial points of our construction is to deal with the distance of $\mathcal{T}$ to being a Lagrangian manifold as a specific error, independent of e. The way in which we address the parameterization method for Lagrangian tori is supported by the basic results stated in Propositions 3.1 and 3.2.

Proposition 3.1 *Let τ be the parameterization of a Lagrangian torus $\mathcal{T}$ of $\mathbb{R}^{2r}$. If $\mathcal{T}$ belongs to an energy level set of the Hamiltonian h, then $\mathcal{T}$ is an invariant torus of h. If in addition τ verifies $\omega = N^{-1}(D\tau)^\top J\nabla h(\tau)$, for some constant vector $\omega \in \mathbb{R}^r$, then τ is a solution of the invariance Eq.* (3).

Proof of Proposition 3.1 We have that $\Omega = 0$ and that the columns of $JD\tau N^{-1}$ and $D\tau$ form a symplectic basis at any point of $\mathcal{T}$. Then, we can write the Hamiltonian vector field of h as follows:

$$J\nabla h(\tau) = D\tau\, a + JD\tau\, b, \tag{25}$$

for some vector functions $a, b : \mathbb{T}^r \to \mathbb{R}^r$. If we multiply (25) by $(D\tau)^\top$ and $(D\tau)^\top J$, then we obtain

$$(D\tau)^\top J\nabla h(\tau) = Na, \qquad (D\tau)^\top \nabla h(\tau) = Nb.$$

Since $\mathcal{T}$ belongs to an energy level set of h, we have that $h(\tau) = h_0$ on $\mathbb{T}^r$, for some constant h_0. By taking partial derivatives with respect to θ, we obtain that $Dh(\tau)D\tau = 0$ and, equivalently, that $(D\tau)^\top \nabla h(\tau) = 0$. This equality means that $b = 0$. Consequently, we can rewrite (25) as

$$J\nabla h(\tau) = D\tau\, N^{-1}(D\tau)^\top J\nabla h(\tau). \tag{26}$$

Formula (26) implies that the correspondence $\theta \in \mathbb{T}^r \mapsto J\nabla h(\tau(\theta))$ defines a tangent vector field on $\mathcal{T}$ (i.e. the vector $J\nabla h(\tau)$ is a combination of the columns of $D\tau$). Hence, $\mathcal{T}$ is an invariant torus. Finally, since we have that $L_\omega \tau = D\tau\, \omega$ and that the rank of the matrix $D\tau$ is maximum at any point, the invariance Eq. (3) is equivalent to the condition $\omega = N^{-1}(D\tau)^\top J\nabla h(\tau)$. □

Proposition 3.1 means that is τ is the parameterization of the torus $\mathcal{T}$, then τ is solution of the invariance Eq. (3) iff it is solution of the following set of nonlinear equations on $\mathbb{T}^r$:

$$(\mathrm{eq}_1)\; h(\tau) = \langle h(\tau)\rangle_\theta, \qquad (\mathrm{eq}_2)\; \Omega = 0, \qquad (\mathrm{eq}_3)\; N^{-1}(D\tau)^\top J\nabla h(\tau) = \omega. \tag{27}$$

We note that the result of Proposition 3.1 holds even if the components of ω are not rationally independent. The same comment applies to Proposition 3.2 below.

Equations in (27) offer a way to address the computation of the parameterization τ which is not based on the direct verification of the invariance condition. However, using equation $\Omega = 0$ to impose that $\mathcal{T}$ is a Lagrangian manifold, we still have the impression that this is not a good way to obtain the desired estimates. The most obvious reason for this assertion is that, since Ω is an anti-symmetric matrix, the equation $\Omega = 0$ implies $r(r-1)/2$ equations for the $2r$-dimensional vector function τ. This means that the condition $\Omega = 0$ involves many dependencies, which are evident at the limit, but that perhaps are not easy to manage iteratively. See also Eq. (30) and comments below. Therefore, our goal is to replace (eq_2) in (27) by an appropriate vectorial equation involving the Lagrangian character of $\mathcal{T}$.

Proposition 3.2 *Let τ be the parameterization of a r-dimensional torus $\mathcal{T}$ of $\mathbb{R}^{2r}$. We suppose that the columns of the matrices $D\tau$ and $JD\tau$ form a basis of $\mathbb{R}^{2r}$ at any point of the torus. We also suppose that $\mathcal{T}$ belongs to an energy level set of h and that τ verifies $\omega = N^{-1}(D\tau)^\top J\nabla h(\tau)$, for some constant vector $\omega \in \mathbb{R}^r$. Then, τ is a solution of the invariance Eq. (3) iff $\Omega\,\omega = 0$.*

Proof of Proposition 3.2 We consider formula (17) for τ, and we discuss the invariance condition $e = 0$. Since the columns of $D\tau$ and $JD\tau$ form a basis at any point, the equation $e = 0$ is equivalent to the set of two equations $(D\tau)^\top e = 0$ and $(D\tau)^\top Je = 0$. If we multiply (17) by $(D\tau)^\top$ and $(D\tau)^\top J$, then we obtain the relations:[1]

$$N\omega = (D\tau)^\top J\nabla h(\tau) + (D\tau)^\top e, \qquad \Omega\,\omega = -(D\tau)^\top \nabla h(\tau) - (D\tau)^\top Je. \tag{28}$$

[1] We note that if $a = a(\theta)$ is a scalar-valued or a vector-valued function, then $L_\omega a = Da\,\omega$.

By using that $\omega = N^{-1}(D\tau)^\top J\nabla h(\tau)$ and that h is constant along $\mathcal{T}$, equations in (28) are equivalent to $(D\tau)^\top e = 0$ and $\Omega\,\omega = -(D\tau)^\top Je$. From these expressions, the result is straightforward. □

Proposition 3.2 implies that we can seek the parameterization τ as solution of the following equations:

$$(\mathrm{eq}_1)\, h(\tau) = \langle h(\tau)\rangle_\theta, \qquad (\mathrm{eq}_2)\, \Omega\,\omega = 0, \qquad (\mathrm{eq}_3)\, N^{-1}(D\tau)^\top J\nabla h(\tau) = \omega. \tag{29}$$

We observe that, since equations in (29) altogether mean that $\mathcal{T} = \tau(\mathbb{T}^r)$ is an invariant torus of h carrying quasi-periodic motion, then the condition $\Omega = 0$ is straightforward from them. However, the condition $\Omega\,\omega = 0$ alone means that $\Omega = 0$. This assertion is consequence of the following relation, which constitutes one of the basic ingredients of the proof of Theorem 2.6 [as a substitute of Eq. (21)]:

$$L_\omega\Omega = D(\Omega\omega) - (D(\Omega\,\omega))^\top. \tag{30}$$

Expression above also supports our (heuristic) claim that we may obtain better estimates by dealing with equation $\Omega\,\omega = 0$ rather than by dealing with equation $\Omega = 0$. Indeed, Eq. (30) means that Ω is naturally related to $\Omega\,\omega$ through the resolution of a small divisor equation. To derive (30), we consider the following homotopy formula (e.g. Weinstein 1979):

$$L_X\varpi = i_X(d\varpi) + d(i_X\varpi). \tag{31}$$

We take as X the constant vector field defined by the frequency vector ω, and we compute the exterior derivative of (31). Then, we obtain:

$$L_\omega(d\varpi) = d(i_\omega(d\varpi)). \tag{32}$$

We apply (32) to the one-form $\varpi = \tau^*(-y\,dx)$. Hence, formula (30) follows from the fact that the coordinate representation of $d\varpi = \tau^*(dx \wedge dy)$ is given by the antisymmetric matrix Ω and that the coordinate representation of $i_\omega(d\varpi)$ is given by the vector $(\omega^\top\Omega)^\top = -\Omega\,\omega$. As a matter of fact, we observe that the direct application of formula (31) to X and ϖ as above shows the following relation between $\Omega\,\omega$ and the vector α giving the coordinate representation of $\tau^*(-ydx)$ (see Definition 2.1):

$$L_\omega\alpha = -\Omega\,\omega + \nabla(\langle\omega, \alpha\rangle). \tag{33}$$

Although formula (33) is not used throughout the paper, we observe that it means that we may consider replacing in (29) equation $\Omega\,\omega = 0$ by the equivalent one $L_\omega\alpha = \nabla(\langle\omega, \alpha\rangle)$.

3.3 Formal Description of the Modified Method

We perform the formal construction of the quasi-Newton method used to prove Theorem 2.6. Given the parameterization τ of a quasi-torus of h, we introduce a new parameterization $\tau^{(1)}$ by approximately solving the linearization of equations (eq$_1$)-(eq$_3$) in (29) around τ. The error functions for τ are:

$$e_1 = h(\tau) - \langle h(\tau)\rangle_\theta, \qquad e_2 = \Omega\,\omega, \qquad e_3 = N^{-1}(D\tau)^\top J\nabla h(\tau) - \omega. \tag{34}$$

Remark 3.3 We note that $\langle e_1\rangle_\theta = 0$ and that $\langle e_2\rangle_\theta = 0$. In particular, equation $e_1 = 0$ is equivalent to $\nabla e_1 = 0$. To apply Proposition 3.2, besides the conditions $e_j = 0$, $j = 1, 2, 3$, we also need to ensure that the columns of the matrices $D\tau$ and $JD\tau$ form a basis at any point. If we define the matrix $\mathcal{N}$ as

$$\mathcal{N} = N\,(\mathrm{Id}_r + N^{-1}\,\Omega\,N^{-1}\,\Omega), \tag{35}$$

then, it is not difficult to realize that this last assertion is equivalent to the fact that $\det\mathcal{N} \neq 0$ at any point. For details, see Eqs. (51) and (54) that express e in terms of e_1, e_2, and e_3. Since $\Omega = 0$ if the torus is invariant, a natural condition to ensure that $\mathcal{N}$ is invertible is that $\|N^{-1}\,\Omega\|_\rho < 1$, for some $\rho > 0$. Finally, we observe that we also can relate each e_j to e as:

$$e_1 = L_\omega^{-1}\langle\nabla h(\tau), e\rangle, \qquad e_2 = -\nabla e_1 + (D\tau)^\top Je, \qquad e_3 = -N^{-1}(D\tau)^\top e. \tag{36}$$

To obtain the first one, compute $L_\omega e_1$ and use (17). For the other two, see Eqs. (52) and (53).

Another crucial aspect of the presented construction is that we seek $\tau^{(1)}$ in the following form:

$$\tau^{(1)}(\theta) = \bar{\tau}^{(1)}(\theta + b(\theta)), \tag{37}$$

where

$$\bar{\tau}^{(1)} = \tau + \Delta\tau, \qquad \Delta\tau = JD\tau\,N^{-1}\,a, \qquad \langle b\rangle_\theta = 0, \tag{38}$$

for some r-dimensional (small) vector functions $a(\theta)$ and $b(\theta)$ to be determined (the condition $\langle b\rangle_\theta = 0$ ensures the uniqueness of solution for b). The reason of considering this expression for $\tau^{(1)}$ is our guess that we may obtain better estimates by using formulas as compact as possible for the involved object. This idea is motivated by the already mentioned approach of Villanueva (2008) to Kolmogorov's theorem. Since a and b are small, the first-order approximation of $\tau^{(1)}$ in (37) is $\tau^{(1)} \approx \tau + JD\tau N^{-1}\,a + D\tau\,b$, which coincides with the expression considered in de la Llave et al. (2005) [see (18)]. Hence, introducing the contribution of b as a composition in (37) only means a small correction on $\tau^{(1)}$. However, in this way we avoid some tricky terms on the error functions for $\tau^{(1)}$ that, perhaps, do not behave appropriately.

Synthetically, the effect of a in formula (38) is to make quadratic both the errors on the energy and errors on the symplectic geometry of the torus $\bar{\tau}^{(1)}$. Hence, the vector function a enters through equations $e_1 = 0$ and $e_2 = 0$. The vector function b means a reparameterization of the torus parameterized by $\bar{\tau}^{(1)}$. So, it only enters through the equation $e_3 = 0$, that is the one responsible of the inner dynamics of the torus.

For further discussions, we also introduce

$$N^{(1)} = (D\tau^{(1)})^\top D\tau^{(1)}, \qquad \Omega^{(1)} = (D\tau^{(1)})^\top J D\tau^{(1)},$$

and the corresponding error functions for $\tau^{(1)}$:

$$e_1^{(1)} = h(\tau^{(1)}) - \langle h(\tau^{(1)}) \rangle_\theta, \;\; e_2^{(1)} = \Omega^{(1)}\,\omega, \;\; e_3^{(1)} = (N^{(1)})^{-1}(D\tau^{(1)})^\top J \nabla h(\tau^{(1)}) - \omega. \tag{39}$$

By linearization around the approximate parameterization τ, we obtain the following expressions:

$$h(\tau^{(1)}) = \langle h(\tau) \rangle_\theta + e_1 - \langle \omega, a \rangle + \mathcal{O}_2, \qquad e_1^{(1)} = e_1 - \langle \omega, \widetilde{a} \rangle + \mathcal{O}_2, \tag{40}$$

$$e_2^{(1)} = e_2 - L_\omega a + \nabla(\langle \omega, a \rangle) + \mathcal{O}_2, \qquad e_3^{(1)} = e_3 - L_\omega b + S\,a + \mathcal{O}_2, \tag{41}$$

$\widetilde{a} = a - \langle a \rangle_\theta$, the (symmetric) matrix S is defined as in (11), and $\mathcal{O}_2$ synthetically denotes the higher-order terms. Computations leading to expressions (40) and (41) are not difficult but involve some cancellations that are not obvious at first glance. We leave the details to Sect. 4. We only emphasize that to obtain these expressions we do not use (explicitly) the automatic reducibility of Lagrangian tori.

Inspection of the expressions in (40) makes it clear that, if we want $e^{(1)}$ to be a quadratic error, then we must select a such that $\langle \omega, \widetilde{a} \rangle = e_1$. There are infinitely many solutions for $\widetilde{a}$ of this equation, but there is only one of the form $\widetilde{a} = \nabla g$, for some scalar function g defined on $\mathbb{T}^r$. Specifically, g verifies:

$$L_\omega g = e_1, \qquad \langle g \rangle_\theta = 0,$$

where the normalizing condition $\langle g \rangle_\theta = 0$ ensures the uniqueness of solution for g. Then, every solution for a of equation $\langle \omega, \widetilde{a} \rangle = e_1$ can be written in the following form:

$$a = \widetilde{a} + \xi, \qquad \widetilde{a} = \nabla g + d, \tag{42}$$

where $\xi = \langle a \rangle_\theta \in \mathbb{R}^r$ is free to take any value and the vector function d is free among those verifying:

$$\langle d \rangle_\theta = 0, \qquad \langle \omega, d \rangle = 0, \tag{43}$$

If we introduce expression for a in (42) in for formula for $e_2^{(1)}$ in (41), then we obtain

$$e_2^{(1)} = e_2 - L_\omega(\nabla g + d) + \nabla(L_\omega g) + \mathcal{O}_2 = e_2 - L_\omega d + \mathcal{O}_2.$$

Hence, to obtain a quadratic expression for $e_2^{(1)}$, we should select d as the solution of the following equation:

$$L_\omega d = e_2, \qquad \langle d \rangle_\theta = 0. \tag{44}$$

By using the definition of e_2 in (34) and the anti-symmetric character of Ω, we have that d verifies:

$$L_\omega(\langle \omega, d \rangle) = \langle \omega, L_\omega d \rangle = \langle \omega, e_2 \rangle = \langle \omega, \Omega\,\omega \rangle = 0.$$

Since the components of the vector ω are rationally independent, equation $L_\omega(\langle \omega, d\rangle) = 0$ means that $\langle \omega, d\rangle$ is a constant function. The normalization $\langle d\rangle_\theta = 0$ implies that $\langle \omega, d\rangle = 0$. Hence, the solution for d of (44) is compatible with conditions in (43). Finally, to achieve a quadratic error $e_3^{(1)}$ in (41), we should define b as follows:

$$b = L_\omega^{-1}(S\,a + e_3). \tag{45}$$

To do this, we need that $\langle Sa + e_3\rangle_\theta = 0$. We can get rid of this average by selecting the following value for ξ in (42):

$$\xi = -(\langle S\rangle_\theta)^{-1}\,(\langle S\,\widetilde{a}\rangle_\theta + \langle e_3\rangle_\theta). \tag{46}$$

As a summary of the above computations, we have established the following expression for a:

$$\widetilde{a} = L_\omega^{-1}(\nabla e_1 + e_2), \qquad a = \widetilde{a} + \xi, \tag{47}$$

with ξ given by (46). Moreover, the vector function a verifies:

$$\langle \omega, a\rangle = e_1 + \langle \omega, \xi\rangle. \tag{48}$$

From now on, we focus on formulas in (45), (46), (47), and (48). Indeed, the role of the functions g and d introduced above is to ensure that the construction is well defined, but their explicit computation is not required to implement the method.

Remark 3.4 Let us summarize the specific steps to be followed to compute the new parameterization $\tau^{(1)}$ from τ. This recap can be useful for people willing to carry out the numerical implementation of the method. Given τ, firstly we compute $D\tau$, $h(\tau)$, $\nabla h(\tau)$, $D^2h(\tau)$, N, N^{-1}, Ω, S, $\langle S\rangle_\theta$, $\{e_j\}_{j=1}^3$, and ∇e_1 [see Eqs. (2), (11), and (34)]. Next, we compute $\widetilde{a}$, ξ, a, and b by the formulas displayed in (45), (46), and (47). Finally, we compute $\Delta\tau$, $\bar{\tau}^{(1)}$, and $\tau^{(1)}$ by the expressions (37) and (38). From the numerical viewpoint, we approximate each of these objects by a trigonometric polynomial of $\theta \in \mathbb{T}^r$. We note that these computations involve several compositions of functions (as well as products of matrices and the computation of N^{-1}). Since we are dealing with expressions which are 2π-periodic in all the variables, it is not necessary to use an algebraic manipulator to carry out this scheme. We just need to evaluate each of these expressions on a suitable r-dimensional grid of points $\theta \in \mathbb{T}^r$ and then performing a FFT. This approach in terms of FFT is one of the advantages of parameterization methods in KAM theory and remain a valid stratgey to perform a CAP. We refer to Calleja and de la Llave (2010), Haro et al. (2016), Huguet et al. (2012) for details.

4 Proof of Theorem 2.6

We proceed by the iterative application of the quasi-Newton method we have developed in Sect. 3.3. In this way, we introduce a sequence $\{\tau^{(n)}\}_{n\geq 0}$ of parameterizations, with $\tau^{(0)} = \tau$, and we show that $\tau^* = \lim_{n\to\infty} \tau^{(n)}$. According to Proposition 3.2, we should verify that $e_1^* = 0$ and that $e_2^* = e_3^* = 0$, where these error functions are defined as in Eq. (34), but now in terms of τ^* instead of τ. As noted in Remark 3.3, we can replace the condition $e_1^* = 0$ by $\nabla e_1^* = 0$. We also need to verify that the columns of the matrices $D\tau^*$ and $JD\tau^*$ form a basis along the torus. This condition is immediate from the proof.

Explicit expressions of one step of the method We consider the initial parameterization τ, the associated matrices N, Ω, S, and $\mathcal{N}$ defined in Eqs. (2), (11), and (35), as well as the error functions e_1, e_2, and e_3 defined in (34). Some useful expressions that follow from (34) are:

$$De_1 = Dh(\tau)\, D\tau, \qquad \nabla e_1 = (D\tau)^\top \nabla h(\tau), \qquad Dh(\tau) J D\tau = -(\omega^\top + e_3^\top)\, N. \tag{49}$$

Key hypotheses for the upcoming computations are that $\det(\langle S\rangle_\theta) \neq 0$ and that N and $\mathcal{N}$ are invertible matrix functions (see Remark 3.3). After application of the method, the new parameterization $\tau^{(1)}$ is defined by (37) and (38), where a, ξ, and b are given by Eqs. (45), (46), (47) [see also (48)]. Our aim is to provide explicit formulas for the (new) errors $e_1^{(1)}$, $e_2^{(1)}$, and $e_3^{(1)}$, defined as in (34), but now in terms of $\tau^{(1)}$ instead of τ. For this purpose, we also denote by $N^{(1)}$, $\Omega^{(1)}$, and $S^{(1)}$ the expressions obtained by computing N, Ω, and S in terms of $\tau^{(1)}$. Further, we introduce the auxiliary expressions $\bar{N}^{(1)}$, $\bar{\Omega}^{(1)}$, $\bar{S}^{(1)}$, $\bar{e}_2^{(1)}$, and $\bar{e}_3^{(1)}$, which are defined like those above, but now in terms of $\bar{\tau}^{(1)}$ [see (38)]. In the following computations, we are only concerned with the formulas of $e_1^{(1)}$, $e_2^{(1)}$, and $e_3^{(1)}$, but we do not discuss the conditions guaranteeing that these formal expressions are well defined for the initial τ.

Remark 4.1 Computation of $\{e_j^{(1)}\}_{j=1}^3$ is performed by expanding them up to first order around τ (plus a quadratic error). By means of some cancellations, due to the cohomological equations, we show that the new errors have (almost) quadratic size with respect to $\{e_j^{(1)}\}_{j=1}^3$. The most technical aspects of the proof appear when we write some terms in a way allowing to check that they behave appropriately with respect to γ and δ (see the second part of the proof for the bounds).

The key points of the proof, those leading to the desired estimates, are the two first steps below [see Eqs. (50) and (51)]. In the proof of the original parameterization theorem of de la Llave et al. (2005), the invariance error e is the (only) source of error for τ and Ω is related to e through Eq. (21). Moreover, in de la Llave et al. (2005) a crucial role is played by the automatic quasi-reducibility, which means controlling the size of the matrices E_1 and E_2 of Eq. (20) in terms of the sizes of e and Ω. As noted before, automatic quasi-reducibility is not explicitly involved in the next presentation.

Firstly, we consider the invariance error e of τ [see (17)] and the matrix representation Ω of the pullback by τ of the symplectic form. Both expressions appear explicitly in some computations, so we need to relate them to e_1, e_2, and e_3. The matrix Ω can be written in terms of e_2 through formula (30):

$$\Omega = L_\omega^{-1}(De_2 - (De_2)^\top). \tag{50}$$

To compute e, we use that the columns of $D\tau$ and $JD\tau$ form a basis at any point. As noted in Remark 3.3, this assertion is equivalent to the invertibility of $\mathcal{N}$. On its turn, the invertibility of $\mathcal{N}$ can be guaranteed by controlling the size of the matrix $N^{-1}\Omega$. Then, we express e as

$$e = D\tau\,\alpha + JD\tau\,\beta, \tag{51}$$

for some vector functions $\alpha(\theta)$ and $\beta(\theta)$. To determine α and β, we multiply both sides of Eq. (51) by $(D\tau)^\top J$ and $N^{-1}(D\tau)^\top$, respectively, thus obtaining the following relations:

$$\Omega\alpha - N\beta = (D\tau)^\top Je, \qquad \alpha + N^{-1}\Omega\beta = N^{-1}(D\tau)^\top e.$$

By performing the same multiplications on Eq. (17), we obtain

$$(D\tau)^\top Je = \Omega\,\omega + (Dh(\tau)D\tau)^\top = e_2 + \nabla e_1, \tag{52}$$

$$N^{-1}(D\tau)^\top e = \omega - N^{-1}(D\tau)^\top J\nabla h(\tau) = -e_3. \tag{53}$$

After some computations, we end up obtaining the next formulas:

$$\alpha = \mathcal{N}^{-1}\left(\Omega N^{-1}(e_2 + \nabla e_1) - N\,e_3\right), \qquad \beta = -\mathcal{N}^{-1}(e_2 + \nabla e_1 + \Omega\,e_3). \tag{54}$$

To compute $e_1^{(1)}$, we use Taylor's formula to expand $h(\bar\tau^{(1)})$ as follows:

$$\begin{aligned} h(\bar\tau^{(1)}) &= h(\tau) + Dh(\tau)JD\tau\,N^{-1}\,a + R_2 = h(\tau) - (\omega^\top + e_3^\top)\,a + R_2 \\ &= \langle h(\tau)\rangle_\theta - \langle\omega, \xi\rangle + \bar E_1^{(1)}, \end{aligned}$$

where[2]

$$\begin{aligned} R_2(\theta) &= (\Delta\tau(\theta))^\top \left[\int_0^1 (1-s)D^2h(\tau(\theta) + s\Delta\tau(\theta))\,\mathrm{d}s\right]\Delta\tau(\theta), \\ \bar E_1^{(1)} &= -\langle e_3, a\rangle + R_2. \end{aligned} \tag{55}$$

After composition with $\theta + b(\theta)$, we obtain

$$e_1^{(1)} = h(\tau^{(1)}) - \langle h(\tau^{(1)})\rangle_\theta = \bar E_1^{(1)}(\theta + b) - \langle \bar E_1^{(1)}(\theta + b)\rangle_\theta. \tag{56}$$

[2] We use that $f(a+x) - f(a) - Df(a)x = \int_0^1 \frac{d}{\mathrm{d}s}[F(s)]\mathrm{d}s = x^\top\left[\int_0^1 (1-s)D^2 f(a+sx)\,\mathrm{d}s\right]x$, where $F(s) = f(a+sx) + (1-s)Df(a+sx)x$.

To display $\bar{e}_2^{(1)}$, we introduce the auxiliary vector function $c = N^{-1}\, a$, so that

$$\Delta\tau = J D\tau N^{-1}\, a = J D\tau\, c = J \sum_{j=1}^{r} c_j\, \partial_{\theta_j} \tau,$$
$$D(\Delta\tau) = J D\tau\, Dc + J \sum_{j=1}^{r} c_j\, \partial_{\theta_j} (D\tau). \tag{57}$$

Hence,

$$(D\tau)^\top J D(\Delta\tau)\, \omega = -N L_\omega c - \sum_{j=1}^{r} c_j\, (D\tau)^\top \partial_{\theta_j} (D\tau)\, \omega,$$
$$(D(\Delta\tau))^\top J D\tau\, \omega = (Dc)^\top N\omega + \sum_{j=1}^{r} c_j\, \partial_{\theta_j} (D\tau)^\top D\tau\, \omega.$$

From these expressions, we obtain:

$$\begin{aligned}
\bar{e}_2^{(1)} &= \mathbb{A}r\Omega^{(1)}\, \omega = (D\bar{\tau}^{(1)})^\top J\, D\bar{\tau}^{(1)}\, \omega = (D\tau + D(\Delta\tau))^\top J\, (D\tau + D(\Delta\tau))\, \omega \\
&= e_2 + (D\tau)^\top J D(\Delta\tau)\, \omega + (D(\Delta\tau))^\top J D\tau\, \omega + (D(\Delta\tau))^\top J D(\Delta\tau)\, \omega \\
&= e_2 - N L_\omega c + (Dc)^\top N\omega + \sum_{j=1}^{r} c_j \left(\partial_{\theta_j} (D\tau)^\top D\tau - (D\tau)^\top \partial_{\theta_j} (D\tau) \right) \omega \\
&\quad + (D(\Delta\tau))^\top J D(\Delta\tau)\, \omega.
\end{aligned}$$

Next step is to express $\bar{e}_2^{(1)}$ in terms of $a = N\, c$. To do that, we use the following relations:

$$L_\omega a = N L_\omega c + (L_\omega N)\, c, \qquad (L_\omega N)\, c = \sum_{j=1}^{r} c_j\, L_\omega((D\tau)^\top \partial_{\theta_j} \tau),$$
$$a = \sum_{j=1}^{r} c_j\, (D\tau)^\top \partial_{\theta_j} \tau, \qquad Da = N Dc + \sum_{j=1}^{r} c_j\, D((D\tau)^\top \partial_{\theta_j} \tau).$$

From the last formula above, we also obtain:

$$\begin{aligned}
(Dc)^\top N\omega &= (Da)^\top \omega - \sum_{j=1}^{r} c_j \left(D((D\tau)^\top \partial_{\theta_j} \tau) \right)^\top \omega \\
&= (D\langle \omega, a\rangle)^\top - \sum_{j=1}^{r} c_j \left(D((L_\omega \tau)^\top \partial_{\theta_j} \tau) \right)^\top.
\end{aligned}$$

The combination of expressions above allows obtaining the following formula for $\bar{e}_2^{(1)}$:

$$\bar{e}_2^{(1)} = e_2 - L_\omega a + (D\langle \omega, a\rangle)^\top + \sum_{j=1}^{r} c_j V_j + (D(\Delta\tau))^\top J D(\Delta\tau)\,\omega,$$

where, for any $j = 1, \dots, r$, the r-dimensional vector function V_j is given by

$$V_j = L_\omega((D\tau)^\top \partial_{\theta_j}\tau) - \Big(D((L_\omega\tau)^\top \partial_{\theta_j}\tau)\Big)^\top + \partial_{\theta_j}(D\tau)^\top L_\omega\tau - (D\tau)^\top \partial_{\theta_j}(L_\omega\tau).$$

It is not difficult to realize that $V_j = 0$. Explicitly, this assertion follows from the expression

$$D((L_\omega\tau)^\top \partial_{\theta_j}\tau) = (L_\omega\tau)^\top \partial_{\theta_j}(D\tau) + (\partial_{\theta_j}\tau)^\top L_\omega(D\tau).$$

To obtain this formula, we use that if u and v are two vector functions then $D(u^\top v) = u^\top Dv + v^\top Du$. Consequently, we have established the following expression for $\bar{e}_2^{(1)}$:

$$\bar{e}_2^{(1)} = (D(\Delta\tau))^\top J D(\Delta\tau)\,\omega = (D(\Delta\tau))^\top J L_\omega(\Delta\tau). \tag{58}$$

Bounding $\bar{e}_2^{(1)}$ from formula (58) requires bounding $L_\omega(\Delta\tau)$. As shown below in the proof, we obtain a better quantitative estimate for $L_\omega(\Delta\tau)$ if, instead of controlling the Lie derivative $L_\omega(\Delta\tau)$ by Cauchy estimates on the partial derivatives of $\Delta\tau$, we consider the following formula for it:

$$\begin{aligned} L_\omega(\Delta\tau) &= L_\omega(J D\tau\, N^{-1} a) \\ &= J L_\omega(D\tau) N^{-1} a + J D\tau\, L_\omega(N^{-1})\, a + J D\tau N^{-1}(\nabla e_1 + e_2), \end{aligned} \tag{59}$$

where we have computed $L_\omega a$ from Eq. (47). Regarding the Lie derivative $L_\omega(D\tau)$ we consider Eq. (22) and, to control $L_\omega(N^{-1})$, we consider the following expressions:

$$L_\omega(N^{-1}) = -N^{-1}\,(L_\omega N)\,N^{-1}, \qquad L_\omega N = (L_\omega(D\tau))^\top D\tau + (D\tau)^\top L_\omega(D\tau). \tag{60}$$

By using strategy above for controlling $L_\omega(D\tau)$ and $L_\omega(N^{-1})$ [as well as $L_\omega c$ in (72)], we not only avoid performing some extra Cauchy estimates, but we also prevent the constant m of Theorem 2.6 from depending on the size of the components of ω. In order to compute $e_2^{(1)}$ from $\bar{e}_2^{(1)}$, we observe that

$$D\tau^{(1)} = D\bar{\tau}^{(1)}(\theta + b)\,(\mathrm{Id} + Db). \tag{61}$$

Then, by taking also into account Eq. (45) for b, we obtain

$$\begin{aligned} e_2^{(1)} &= \Omega^{(1)}\,\omega = (D\tau^{(1)})^\top J D\tau^{(1)}\,\omega = (\mathrm{Id} + Db)^\top \mathbb{A}r\Omega^{(1)}(\theta + b)\,(\omega + L_\omega b) \\ &= (\mathrm{Id} + Db)^\top (\bar{e}_2^{(1)}(\theta + b) + \mathbb{A}r\Omega^{(1)}(\theta + b)\,(S\,a + e_3)). \end{aligned} \tag{62}$$

In analogy to Eq. (50) for Ω, we can relate $\mathbb{A}r\Omega^{(1)}$ above to $\bar{e}_2^{(1)} = \mathbb{A}r\Omega^{(1)}\,\omega$ through formula (30):

$$\mathbb{A}r\Omega^{(1)} = L_\omega^{-1}\left(D\bar{e}_2^{(1)} - \left(D\bar{e}_2^{(1)}\right)^\top\right). \tag{63}$$

Next, we consider the auxiliary error $\bar{e}_3^{(1)}$:

$$\bar{e}_3^{(1)} = (\bar{N}^{(1)})^{-1}(D\bar{\tau}^{(1)})^\top J\nabla h(\bar{\tau}^{(1)}) - \omega. \tag{64}$$

We point out that we do not claim that $\bar{e}_3^{(1)}$ is a quadratic error [see Eq. (74)]. We only require a quadratic behaviour for $e_3^{(1)}$ [see Eq. (78)]. To control (64), we introduce the following notations:

$$\begin{aligned} \bar{N}^{(1)} &= (D\bar{\tau}^{(1)})^\top D\bar{\tau}^{(1)} = N + \Delta\bar{N} = N\,(\mathrm{Id}_r + N^{-1}\,\Delta\bar{N}),\\ \nabla h(\bar{\tau}^{(1)}) &= \nabla h(\tau) + \bar{R}_1 = \nabla h(\tau) + D^2h(\tau)\Delta\tau + \bar{R}_2, \end{aligned} \tag{65}$$

where[3]

$$\Delta\bar{N} = (D\tau)^\top D(\Delta\tau) + (D(\Delta\tau))^\top D\tau + (D(\Delta\tau))^\top D(\Delta\tau), \tag{66}$$

$$\bar{R}_1(\theta) = \left[\int_0^1 D^2h(\tau(\theta) + s\Delta\tau(\theta))\,ds\right]\Delta\tau(\theta), \tag{67}$$

$$\bar{R}_{2;j}(\theta) = (\Delta\tau(\theta))^\top\left[\int_0^1 (1-s)D^2(\partial_{z_j}h)(\tau(\theta) + s\Delta\tau(\theta))\,\mathrm{d}s\right]\Delta\tau(\theta), \tag{68}$$

for any $j = 1, \ldots, 2r$. Equation (65) implies that the invertibility of $\bar{N}^{(1)}$ is equivalent to the invertibility of $\mathrm{Id}_r + N^{-1}\Delta\bar{N}$ which, on its turn, is guaranteed if the norm of $N^{-1}\Delta\bar{N}$ is smaller than one. We assume that $\mathrm{Id}_r + N^{-1}\Delta\bar{N}$ is invertible and we compute (see definition of e_3 in (34)):

$$\begin{aligned} (\mathrm{Id}_r + N^{-1}\Delta\bar{N})\,\bar{e}_3^{(1)} &= N^{-1}(D\tau + D(\Delta\tau))^\top J\nabla h(\bar{\tau}^{(1)}) - (\mathrm{Id} + N^{-1}\Delta\bar{N})\,\omega\\ &= e_3 + N^{-1}\left((D\tau)^\top JD^2h(\tau)\Delta\tau + (D(\Delta\tau))^\top J\nabla h(\tau) - \Delta\bar{N}\,\omega\right)\\ &\quad + N^{-1}\left((D\tau)^\top J\bar{R}_2 + (D(\Delta\tau))^\top J\bar{R}_1\right). \end{aligned} \tag{69}$$

To control the rightmost part of Eq. (69), we discuss more precisely $\Delta\bar{N}\,\omega$ and $(D(\Delta\tau))^\top J\nabla h(\tau)$. Regarding $\Delta\bar{N}\omega$, we perform the following auxiliary computations [see (57)]:

[3] We have used that $f(a+x) - f(a) = \int_0^1 \frac{d}{ds}[F(s)]\,ds = \left[\int_0^1 Df(a+sx)\,ds\right]x$, where $F(s) = f(a+sx)$.

$$(D\tau)^\top D(\Delta\tau)\,\omega = \Omega\, L_\omega c + \sum_{j=1}^{r} c_j\, (D\tau)^\top J\, \partial_{\theta_j}(D\tau)\,\omega,$$

$$(D(\Delta\tau))^\top D\tau\,\omega = -(Dc)^\top \Omega\,\omega - \sum_{j=1}^{r} c_j\, \partial_{\theta_j}(D\tau)^\top J\, D\tau\,\omega.$$

We apply expressions above to formula (66), and we obtain:

$$\Delta\bar{N}\,\omega = \Omega\, L_\omega c - (Dc)^\top e_2 + W + (D(\Delta\tau))^\top L_\omega(\Delta\tau), \tag{70}$$

where

$$W = \sum_{j=1}^{r} c_j \left((D\tau)^\top J\, \partial_{\theta_j}(D\tau) - \partial_{\theta_j}(D\tau)^\top J\, D\tau \right) \omega. \tag{71}$$

To simplify Eq. (71), we take partial derivatives with respect θ_j, $j = 1, \ldots, r$, in the definitions of e_2 in (34) and of e in (17) (this latter derivative is given by the column j of De in (22)). Then, we obtain:

$$\begin{aligned} W &= \sum_{j=1}^{r} c_j \left(2(D\tau)^\top J\, \partial_{\theta_j}(D\tau)\,\omega - \partial_{\theta_j} e_2 \right) \\ &= 2(D\tau)^\top J \sum_{j=1}^{r} c_j \left(J\, D^2 h(\tau)\, \partial_{\theta_j}\tau + \partial_{\theta_j} e \right) - (De_2)\, c \\ &= -2(D\tau)^\top D^2 h(\tau)\, D\tau\, c + 2\,(D\tau)^\top J\,(De)\, c - (De_2)\, c. \end{aligned}$$

To control the contributions of $L_\omega(\Delta\tau)$ and $L_\omega c$ to Eq. (70), we use Eq. (59) for $L_\omega(\Delta\tau)$ and we rewrite $L_\omega c$ as follows [see Eq. (47)]:

$$L_\omega c = L_\omega(N^{-1}\, a) = L_\omega(N^{-1})\, a + N^{-1}(\nabla e_1 + e_2). \tag{72}$$

Regarding $(D(\Delta\tau))^\top J \nabla h(\tau)$, once again we use Eq. (57). We obtain:

$$(D(\Delta\tau))^\top J \nabla h(\tau) = (Dc)^\top (D\tau)^\top \nabla h(\tau) + \sum_{j=1}^{r} c_j\, \partial_{\theta_j}(D\tau)^\top \nabla h(\tau).$$

To simplify expressions above, we consider the formula for ∇e_1 given in Eq. (49). In particular, if we take partial derivatives with respect to θ_j of this formula, for any $j = 1, \ldots, r$, then we obtain:

$$\partial_{\theta_j}(D\tau)^\top \nabla h(\tau) + (D\tau)^\top D^2 h(\tau)\, \partial_{\theta_j}\tau = \partial_{\theta_j}(\nabla e_1).$$

Consequently,

$$
\begin{aligned}
(D(\Delta\tau))^\top J\nabla h(\tau) &= (Dc)^\top \nabla e_1 + \sum_{j=1}^{r} c_j \, (\partial_{\theta_j}(\nabla e_1) - (D\tau)^\top D^2 h(\tau)\, \partial_{\theta_j}\tau) \\
&= (Dc)^\top \nabla e_1 - (D\tau)^\top D^2 h(\tau) D\tau \, c + (D^2 e_1)\, c.
\end{aligned}
$$

We use computations above to rewrite (69) as follows [see formula (38) for $\Delta\tau$ and recall $c = N^{-1}a$]:

$$
(\mathrm{Id}_r + N^{-1}\Delta\bar{N})\, \bar{e}_3^{(1)} = e_3 + S\, a + \bar{E}_2^{(1)},
$$

where S is the symmetric matrix defined in (11) and

$$
\begin{aligned}
\bar{E}_2^{(1)} = N^{-1} \Big(&(Dc)^\top (\nabla e_1 + e_2) - \Omega\, L_\omega c + ((D^2 e_1) + De_2 - 2\,(D\tau)^\top J\,(De))\, c \\
&-(D(\Delta\tau))^\top L_\omega(\Delta\tau) + (D\tau)^\top J \bar{R}_2 + (D(\Delta\tau))^\top J \bar{R}_1 \Big). \qquad (73)
\end{aligned}
$$

From these expressions, we obtain

$$
\bar{e}_3^{(1)} = (\mathrm{Id}_r + N^{-1}\Delta\bar{N})^{-1}(e_3 + S\, a + \bar{E}_2^{(1)}) = e_3 + S\, a + \bar{E}_3^{(1)}, \qquad (74)
$$

where

$$
\bar{E}_3^{(1)} = \Big((\mathrm{Id}_r + N^{-1}\Delta\bar{N})^{-1} - \mathrm{Id}_r \Big)\, (e_3 + S\, a) + (\mathrm{Id}_r + N^{-1}\Delta\bar{N})^{-1}\, \bar{E}_2^{(1)}. \qquad (75)
$$

To display $e_3^{(1)}$, we consider the following computations [see (37), (39), (45), (61), and (64)]:

$$
\begin{aligned}
(\mathrm{Id}_r + Db)\, e_3^{(1)} &= (\mathrm{Id}_r + Db) \Big((N^{(1)})^{-1} (D\tau^{(1)})^\top J \nabla h(\tau^{(1)}) - \omega \Big) \\
&= (\bar{N}^{(1)}(\theta + b))^{-1} (D\bar{\tau}^{(1)}(\theta + b))^\top J \nabla h(\bar{\tau}^{(1)}(\theta + b)) - (\mathrm{Id}_r + Db)\, \omega \\
&= \bar{e}_3^{(1)}(\theta + b) - L_\omega b = \bar{e}_3^{(1)}(\theta + b) - e_3 - S\, a, \qquad (76)
\end{aligned}
$$

where we also used that

$$
N^{(1)} = (D\tau^{(1)})^\top D\tau^{(1)} = (\mathrm{Id}_r + Db)^\top\, \bar{N}^{(1)}(\theta + b)\, (\mathrm{Id}_r + Db). \qquad (77)
$$

If we apply Eqs. (74) to (76), then we obtain the following formula por $e_3^{(1)}$

$$
e_3^{(1)} = (\mathrm{Id}_r + Db)^{-1}\, \bar{E}_4^{(1)}, \qquad (78)
$$

where, in order to make clear the quadratic behaviour of $e_3^{(1)}$, we write $\bar{E}_4^{(1)}$ as follows:

$$
\bar{E}_4^{(1)} = \bar{E}_3^{(1)}(\theta + b) + e_3(\theta + b) - e_3 + S(\theta + b)\, (a(\theta + b) - a) + (S(\theta + b) - S)\, a. \qquad (79)
$$

Indeed, we can write

$$e_3(\theta + b(\theta)) - e_3(\theta) = \left(\int_0^1 De_3(\theta + s\, b(\theta))\, ds \right) b(\theta), \tag{80}$$

as well as analogous expressions for $a(\theta + b) - a$ and for any column of the matrix $S(\theta + b) - S$.

Bounding one step of the method We derive quantitative estimates on the different objects involved on procedure above. The only required tools are Lemma 2.4 (to bound the solutions of the small divisors equations), as well as some basic properties of the norm $\|\cdot\|_\rho$ [see Eq. (6) and Remark 2.5 for the definition]. Explicitly, if f is an analytic function of $\mathbb{T}^r$ (scalar-valued, vector-valued or matrix-valued), then we have the following bounds for the average of f and for the derivatives of f (Cauchy estimates):

$$|\langle f \rangle_\theta| \leq \|f\|_\rho, \qquad \|\widetilde{f}\|_\rho \leq 2\|f\|_\rho, \qquad \|\partial_{\theta_j} f\|_{\rho-\delta} \leq \frac{\|f\|_\rho}{\delta}, \quad j = 1, \ldots, r,$$

where $\widetilde{f} = f - \langle f \rangle_\theta$. Moreover, we control the inverse of a matrix that is close to the identity by Neumann's series. Indeed, if $A \in \mathbb{M}_{r\times r}(\mathbb{C})$ verifies $|A| < 1$, then the matrix $\mathrm{Id}_r + A$ verifies:

$$|(\mathrm{Id}_r + A)^{-1}| \leq \frac{1}{1 - |A|}, \qquad |(\mathrm{Id}_r + A)^{-1} - \mathrm{Id}| \leq \frac{|A|}{1 - |A|}.$$

If A is an analytic matrix function of $\mathbb{T}^r$, then expressions above also hold for the norm $\|\cdot\|_\rho$.

Let τ be the parameterization of the quasi-torus of the statement of Theorem 2.6. Our aim is to construct a constant m for which the theorem holds. To achieve this purpose, during the next sequence of bounds the value of m is redefined recursively to meet a finite number of conditions. The *last* value of m obtained is the one of the statement. We use the constants of the statement of Theorem 2.6 to introduce the following normalized error:

$$\mu = \mu_3 + \frac{\mu_1 + \mu_2}{\gamma \delta^\nu}.$$

To guarantee that we can apply the quasi-Newton method to the parameterization τ and to provide quantiative estimates on the involved objects, we suppose that the following bound holds for all the values of m (i.e. for the last one) of this finite sequence of constants [see (13) and compare with condition (16)]:

$$m \frac{\mu}{\gamma \delta^{\nu+1}} < \min \left\{ \frac{1}{2}, \Theta \right\}. \tag{81}$$

Firstly, we use Lemma 2.4 and the Cauchy estimates to bound the matrix Ω from (50):

$$\|\Omega\|_{\rho-\delta}=\|L_\omega^{-1}(De_2-(De_2)^\top)\|_{\rho-\delta} \le \sigma \frac{\|De_2-(De_2)^\top\|_\rho}{\gamma\delta^\nu} \le \frac{2r\sigma\|e_2\|_\rho}{\gamma\delta^{\nu+1}} \le m\frac{\mu}{\delta}.$$

By redefining m, we also have that $\|N^{-1}\Omega\|_{\rho-\delta} \le m\mu/\delta$, so (81) implies that $\|N^{-1}\Omega\|_{\rho-\delta} \le 1/\sqrt{2}$. Hence, by Neumann's series, the matrix $\mathcal{N}$ defined in (35) is non-singular and verifies:

$$\|\mathcal{N}^{-1}\|_{\rho-\delta} \le \frac{\|N^{-1}\|_\rho}{1-\|N^{-1}\Omega\|^2_{\rho-\delta}} \le 2m_4 \le m.$$

By Remark 3.3, this means that the columns of the matrix functions $D\tau$ and $JD\tau$ form a basis at any point of the set $\Delta(\rho-\delta)$. This fact allows using formulas (51) and (54) to obtain:

$$\|\alpha\|_{\rho-\delta} \le m\mu, \quad \|\beta\|_{\rho-\delta} \le m\mu, \quad \|e\|_{\rho-\delta} \le m\mu, \quad \|De\|_{\rho-2\delta} \le m\frac{\mu}{\delta},$$

where we use (81) to ensure that $\mu/\delta \le 1$. From the rightmost part of Eq. (22) and formulas in (60), we also obtain

$$\|L_\omega(D\tau)\|_{\rho-2\delta} \le m, \quad \|L_\omega N\|_{\rho-2\delta} \le m, \quad \|L_\omega(N^{-1})\|_{\rho-2\delta} \le m.$$

From Eqs. (11), (38), (45), (46), (47), and (59), we obtain (we recall that $c = N^{-1}a$):

$$\|S\|_\rho \le m, \quad \|\tilde{a}\|_{\rho-\delta} \le m\mu, \quad |\xi| \le m\mu, \quad \|a\|_{\rho-\delta} \le m\mu, \quad \|b\|_{\rho-2\delta} \le m\frac{\mu}{\gamma\delta^\nu},$$
$$\|c\|_{\rho-\delta} \le m\mu, \quad \|\Delta\tau\|_{\rho-\delta} \le m\mu, \quad \|D(\Delta\tau)\|_{\rho-2\delta} \le m\frac{\mu}{\delta}, \quad \|L_\omega(\Delta\tau)\|_{\rho-2\delta} \le m\mu.$$

As discussed before, the discrepancy between the bounds on $D(\Delta\tau)$ and $L_\omega(\Delta\tau)$ above is due to the fact that the size of $D(\Delta\tau)$ is controlled by Cauchy estimates while for $L_\omega(\Delta\tau)$ we use formula (59).

To guarantee that the new parameterization $\tau^{(1)}(\theta) = \bar{\tau}^{(1)}(\theta+b(\theta))$ is well defined [see Eqs. (37) and (38)], we need to control the domain of definition of this composition of functions. By condition (81) and bounds above, we have that $\|b\|_{\rho-2\delta} \le \delta$. This fact implies that

$$(\cdot + sb(\cdot))(\Delta(\rho-j\delta)) \subset \Delta(\rho-(j-1)\delta), \qquad \forall s \in [0,1], \quad j = 2,3,4, \tag{82}$$

and that this inclusion also holds for the closures of $\Delta(\rho-j\delta)$ and $\Delta(\rho-(j-1)\delta)$. Since we have that $\bar{\tau}^{(1)} = \tau + \Delta\tau$ is well defined on the closure of $\Delta(\rho-\delta)$, we have that $\tau^{(1)}$ is analytic on $\Delta(\rho-2\delta)$. To control $\tau^{(1)} - \tau$, we express it as follows:

$$\begin{aligned}\tau^{(1)}(\theta) - \tau(\theta) &= \tau(\theta + b(\theta)) - \tau(\theta) + \Delta\tau(\theta + b(\theta)) \\ &= \left(\int_0^1 D\tau(\theta + sb(\theta))\, ds\right) b(\theta) + \Delta\tau(\theta + b(\theta)).\end{aligned}$$

Hence, we have

$$\|\tau^{(1)} - \tau\|_{\rho-2\delta} \leq m\frac{\mu}{\gamma\delta^\nu}, \qquad \|D\tau^{(1)} - D\tau\|_{\rho-3\delta} \leq m\frac{\mu}{\gamma\delta^{\nu+1}}.$$

Then, we can control the distance of $\tau^{(1)}(\Delta(\rho-2\delta))$ to the boundary of $\mathcal{U}$ as follows [see Definition (10)]:

$$d_{\tau^{(1)},\rho-2\delta,\mathcal{U}} \geq d_{\tau,\rho,\mathcal{U}} - \|\tau^{(1)} - \tau\|_{\rho-2\delta} \geq d_{\tau,\rho,\mathcal{U}} - m\frac{\mu}{\gamma\delta^\nu} > m_5, \tag{83}$$

where the last inequality follows from condition (81). Bound (83) implies that we can evaluate h and its derivatives on the set $\tau^{(1)}(\Delta(\rho-2\delta))$. By using once again condition (81), we have

$$\|D\tau^{(1)}\|_{\rho-3\delta} \leq \|D\tau\|_\rho + \|D\tau^{(1)} - D\tau\|_{\rho-3\delta} \leq \|D\tau\|_\rho + m\frac{\mu}{\gamma\delta^{\nu+1}} < m_3.$$

To show that the matrix function $N^{(1)}$ is non-singular, we consider the following bounds [see (66)]:

$$\|\Delta\bar{N}\|_{\rho-2\delta} \leq m\frac{\mu}{\delta}, \qquad \|N^{-1}\Delta\bar{N}\|_{\rho-2\delta} \leq m\frac{\mu}{\delta}, \qquad \|Db\|_{\rho-3\delta} \leq m\frac{\mu}{\gamma\delta^{\nu+1}}.$$

By condition (81), we have that $m\mu/\delta \leq 1/2$ and that $m\mu/(\gamma\delta^{\nu+1}) \leq 1/2$. Hence, the matrix functions $\bar{\mathcal{N}} = \mathrm{Id}_r + N^{-1}\Delta\bar{N}$, $\bar{N}^{(1)}$ [see (65)], and $\mathrm{Id} + Db$ are non-singular and verify:

$$\begin{aligned}&\|\bar{\mathcal{N}}^{-1}\|_{\rho-2\delta} \leq 2, \quad \|\bar{\mathcal{N}}^{-1} - \mathrm{Id}\|_{\rho-2\delta} \leq m\frac{\mu}{\delta}, \\ &\|(\bar{N}^{(1)})^{-1}\|_{\rho-2\delta} \leq m, \quad \|(\mathrm{Id} + Db)^{-1}\|_{\rho-3\delta} \leq 2.\end{aligned}$$

From bounds above and formula (77), we conclude that $N^{(1)}$ is non-singular and we derive the straightforward bound $\|(N^{(1)})^{-1}\|_{\rho-3\delta} \leq m$. We note that to provide this latter estimate, we should take into account that the composition $(\bar{N}^{(1)})^{-1}(\theta + b(\theta))$ is well-defined $\forall\theta \in \Delta(\rho-3\delta)$ [see inclusion (82) for $j = 3$]. To bound $(N^{(1)})^{-1}$ more accurately, we consider the following formulas:

$$
\begin{aligned}
N^{-1} - (N^{(1)})^{-1} &= (N^{(1)})^{-1}(N^{(1)} - N)N^{-1} \\
&= (N^{(1)})^{-1} \Big[(\mathrm{Id}_r + Db)^\top \Delta\bar{N}(\theta + b)\,(\mathrm{Id}_r + Db) \\
&\quad + (Db)^\top N(\theta + b) \\
&\quad + (\mathrm{Id}_r + Db)^\top N(\theta + b)\, Db + N(\theta + b) - N \Big] N^{-1},
\end{aligned}
$$

$$
N_{i,j}(\theta + b(\theta)) - N_{i,j}(\theta) = \left(\int_0^1 DN_{i,j}(\theta + sb(\theta))\, ds \right) b(\theta), \quad i, j = 1, \ldots, r.
$$

From these expressions, we derive the following bounds [see condition (81)]:

$$
\begin{aligned}
&\|N(\theta + b) - N\|_{\rho - 3\delta} \le m \frac{\mu}{\gamma \delta^{\nu+1}}, \\
&\|(N^{(1)})^{-1} - N^{-1}\|_{\rho - 3\delta} \le m \frac{\mu}{\gamma \delta^{\nu+1}}, \quad \|(N^{(1)})^{-1}\|_{\rho - 3\delta} < m_4.
\end{aligned}
$$

Next step is to deal with the matrix $\langle S^{(1)} \rangle_\theta$, where [compare with the definition of S in (11)]:

$$
S^{(1)} = (N^{(1)})^{-1} (D\tau^{(1)})^\top \left[J D^2 h(\tau^{(1)}) J + D^2 h(\tau^{(1)}) \right] D\tau^{(1)} (N^{(1)})^{-1}.
$$

The difference $S^{(1)} - S$ can be controlled in terms of bounds on N^{-1}, $(N^{(1)})^{-1}$, $D\tau$, $D\tau^{(1)}$, $D^2h(\tau)$, $D^2h(\tau^{(1)})$, $(N^{(1)})^{-1} - N^{-1}$, $D\tau^{(1)} - D\tau$, and $D^2h(\tau^{(1)}) - D^2h(\tau)$. We observe that hypothesis (8) implies that $\|D^2h(\tau)\|_\rho \le m_1$ and that $\|D^2h(\tau^{(1)})\|_{\rho-3\delta} \le m_1$, since these compositions are well defined on the corresponding domains. Hence, $D^2h(\tau^{(1)}) - D^2h(\tau)$ is the only expression of the list that we have not yet considered. By writing $D^2h(\tau^{(1)}) - D^2h(\tau)$ column by column, we have

$$
\begin{aligned}
&\nabla(\partial_{z_j} h)(\tau^{(1)}) - \nabla(\partial_{z_j} h)(\tau) \\
&\quad = \left(\int_0^1 D^2(\partial_{z_j} h)(\tau + s(\tau^{(1)} - \tau))\, ds \right) (\tau^{(1)} - \tau), \quad j = 1, \ldots, 2r.
\end{aligned}
$$

By means of analogous computations as those performed in (83), we have that $d_{\tau + s(\tau^{(1)} - \tau), \rho - 3\delta, \mathcal{U}} > m_5$, $\forall s \in [0, 1]$. Hence, by using hypothesis (8), we conclude that $\|D^2(\partial_{z_j} h)(\tau + s(\tau^{(1)} - \tau))\|_{\rho - 3\delta} \le m_2$. Consequently,

$$
\|D^2h(\tau^{(1)}) - D^2h(\tau)\|_{\rho - 3\delta} \le m \frac{\mu}{\gamma \delta^\nu}, \qquad \|S^{(1)} - S\|_{\rho - 3\delta} \le m \frac{\mu}{\gamma \delta^{\nu+1}}.
$$

Now, we write:

$$
\langle S^{(1)} \rangle_\theta = \langle S \rangle_\theta \, (\mathrm{Id}_r + (\langle S \rangle_\theta)^{-1} \, (\langle S^{(1)} \rangle_\theta - \langle S \rangle_\theta)),
$$

and we consider the bound

$$|(\langle S\rangle_\theta)^{-1}\,(\langle S^{(1)}\rangle_\theta - \langle S\rangle_\theta)| \leq m_6\, \|S^{(1)} - S\|_{\rho-3\delta} \leq m\frac{\mu}{\gamma\delta^{\nu+1}}.$$

Since $m\mu/(\gamma\delta^{\nu+1}) \leq 1/2$, we have that $\langle S^{(1)}\rangle_\theta$ is non-singular and, by bounding Neumann's series,

$$\begin{aligned}&|(\langle S^{(1)}\rangle_\theta)^{-1} - (\langle S\rangle_\theta)^{-1}| = \\ &\quad |((\mathrm{Id}_r + (\langle S\rangle_\theta)^{-1}\,(\langle S^{(1)}\rangle_\theta - \langle S\rangle_\theta))^{-1} - \mathrm{Id}_r)\,(\langle S\rangle_\theta)^{-1}| \leq m\frac{\mu}{\gamma\delta^{\nu+1}}.\end{aligned}$$

By using once again condition (81), we deduce:

$$|(\langle S^{(1)}\rangle_\theta)^{-1}| \leq |(\langle S\rangle_\theta)^{-1}| + |(\langle S^{(1)}\rangle_\theta)^{-1} - (\langle S\rangle_\theta)^{-1}| \leq |(\langle S\rangle_\theta)^{-1}| + m\frac{\mu}{\gamma\delta^{\nu+1}} < m_6.$$

To finish this part of the proof, we should bound the new errors $\{e_j^{(1)}\}_{j=1}^3$. From (55) and (56), we have:

$$\|R_2\|_{\rho-\delta} \leq m\mu^2 \quad \|\bar{E}_1^{(1)}\|_{\rho-\delta} \leq m\mu^2, \quad \|e_1^{(1)}\|_{\rho-2\delta} \leq m\mu^2, \quad \|\nabla e_1^{(1)}\|_{\rho-3\delta} \leq m\frac{\mu^2}{\delta}.$$

To control R_2, we used that $m\mu < d_{\tau,\rho,\mathcal{U}} - m_5$ [see (81)] to derive the bound $d_{\tau+s\Delta\tau,\rho-\delta,\mathcal{U}} > m_5$, $\forall s \in [0,1]$ [compare with (83)]. Then, we have that $\|D^2h(\tau + s\Delta\tau)\|_{\rho-\delta} \leq m_1$. Moreover, to control the composition $\bar{E}_1^{(1)}(\theta+b)$, we used (82) for $j=2$. From Eqs. (58), (62), and (63), we obtain:

$$\|\bar{e}_2^{(1)}\|_{\rho-2\delta} \leq m\frac{\mu^2}{\delta}, \qquad \|\mathbb{A}r\Omega^{(1)}\|_{\rho-4\delta} \leq m\frac{\mu^2}{\gamma\delta^{\nu+2}}, \qquad \|e_2^{(1)}\|_{\rho-5\delta} \leq m\frac{\mu^2}{\delta},$$

where to control $e_2^{(1)}$ we have used (82) for $j=2,4$ and that $\mu/(\gamma\delta^{\nu+1}) \leq 1$ [see (81)]. To estimate $e_3^{(1)}$ from (78), we should consider Eqs. (67), (68) (for any $j=1,\ldots,2r$), (72), (73), (75), (79), and (80) (as well as analogous expressions for $a(\theta+b)-a$ and $S(\theta+b)-S$). We obtain (involved compositions have been discussed above):

$$\begin{aligned}&\|\bar{R}_1\|_{\rho-\delta} \leq m\mu, \quad \|\bar{R}_2\|_{\rho-\delta} \leq m\mu^2, \quad \|L_\omega c\|_{\rho-2\delta} \leq m\mu,\\ &\|S(\theta+b) - S\|_{\rho-2\delta} \leq m\frac{\mu}{\gamma\delta^{\nu+1}},\\ &\|e_3(\theta+b) - e_3\|_{\rho-2\delta} \leq m\frac{\mu^2}{\gamma\delta^{\nu+1}}, \quad \|a(\theta+b) - a\|_{\rho-3\delta} \leq m\frac{\mu^2}{\gamma\delta^{\nu+1}},\\ &\|\bar{E}_2^{(1)}\|_{\rho-2\delta} \leq m\frac{\mu^2}{\delta}, \quad \|\bar{E}_3^{(1)}\|_{\rho-2\delta} \leq m\frac{\mu^2}{\delta},\\ &\|\bar{E}_3^{(1)}(\theta+b)\|_{\rho-3\delta} \leq m\frac{\mu^2}{\delta}, \quad \|\bar{E}_4^{(1)}\|_{\rho-3\delta} \leq m\frac{\mu^2}{\gamma\delta^{\nu+1}}.\end{aligned}$$

From them, we obtain the estimate

$$\|e_3^{(1)}\|_{\rho-3\delta} \leq m\frac{\mu^2}{\gamma\delta^{\nu+1}}.$$

From now on, we set the value of the constant m as the one for which all previous estimates are fulfilled. We use bounds above as motivation to introduce the following quantities:

$$\rho^{(1)} = \rho - 5\delta, \quad \mu_1^{(1)} = \mu_2^{(1)} = m\frac{\mu^2}{\delta}, \quad \mu_3^{(1)} = m\frac{\mu^2}{\gamma\delta^{\nu+1}}.$$

Hence, at this point we have established the following conclusions. Let us suppose that hypotheses on the statement of Theorem 2.6 hold and that μ verifies (81) for the selected value of δ. Then, inequalities (9), (12), and (15) of the statement still hold if we replace τ, N, S, $\{e_j\}_{j=1}^3$, ρ, and $\{\mu_j\}_{j=1}^3$ by $\tau^{(1)}$, $N^{(1)}$, $S^{(1)}$, $\{e_j^{(1)}\}_{j=1}^3$, $\rho^{(1)}$, and $\{\mu_j^{(1)}\}_{j=1}^3$, respectively. Moreover, if we define:

$$\Lambda = \max\left\{ \|D\tau^{(1)} - D\tau\|_{\rho^{(1)}},\ \|(N^{(1)})^{-1} - N^{-1}\|_{\rho^{(1)}},\ |(\langle S^{(1)}\rangle_\theta)^{-1} - (\langle S\rangle_\theta)^{-1}| \right\}, \tag{84}$$

then we have proved that

$$\|\tau^{(1)} - \tau\|_{\rho^{(1)}} \leq m\frac{\mu}{\gamma\delta^{\nu}}, \qquad \Lambda \leq m\frac{\mu}{\gamma\delta^{\nu+1}}. \tag{85}$$

Iterative application of the method We conclude the proof by showing that the iterative application of the method to the initial parameterization τ converges to τ^*. We introduce the following notations:

$$\tau^{(0)} = \tau, \quad N^{(0)} = N, \quad S^{(0)} = S, \quad \rho^{(0)} = \rho, \quad \delta^{(0)} = \delta, \quad \mu^{(0)} = \mu,$$
$$\mu_j^{(0)} = \mu_j, \quad j = 1, 2, 3.$$

Let us assume that we can iterate n times, $n \geq 1$. Then, we denote by $\tau^{(n)}$, $N^{(n)}$, and $S^{(n)}$, the corresponding expressions after the n-iteration. Moreover, we also introduce the following recurrences (defined in terms of the value of m already fixed):

$$\delta^{(n)} = \frac{\delta^{(n-1)}}{2}, \quad \rho^{(n)} = \rho^{(n-1)} - 5\delta^{(n-1)}, \quad \mu_1^{(n)} = \mu_2^{(n)} = m\frac{(\mu^{(n-1)})^2}{\delta^{(n-1)}},$$
$$\mu_3^{(n)} = m\frac{(\mu^{(n-1)})^2}{\gamma(\delta^{(n-1)})^{\nu+1}},$$
$$\mu^{(n)} = \mu_3^{(n)} + \frac{\mu_1^{(n)} + \mu_2^{(n)}}{\gamma(\delta^{(n)})^{\nu}} = \frac{m(1+2^{\nu+1})}{\gamma\delta^{\nu+1}} 2^{(n-1)(\nu+1)}(\mu^{(n-1)})^2 = 2^{-n(\nu+1)}\chi^{2^n-1}\mu,$$

where

$$\chi = m\frac{2^{\nu+1}(1+2^{\nu+1})}{\gamma\delta^{\nu+1}}\mu. \tag{86}$$

In particular, we have that $\rho^* = \lim_{n\to\infty}\rho^{(n)} = \rho - 10\delta$. According to the notations introduced above and the bounds that we have established for the first step of the process, we conclude that can iterate n times provided that the following bounds hold [see (81)]:

$$m\frac{\mu^{(j)}}{\gamma(\delta^{(j)})^{\nu+1}} < \min\left\{\frac{1}{2}, \Theta^{(j)}\right\}, \qquad j = 0,\ldots,n-1, \tag{87}$$

where $\Theta^{(j)}$ is defined as Θ in (13), by replacing τ, N, S, and ρ, by $\tau^{(j)}$, $N^{(j)}$, $S^{(j)}$, and $\rho^{(j)}$, respectively. In particular $\Theta^{(0)} = \Theta$, so the inequality (87) for $j = 0$ reads like (81). Our aim is to show that, if condition (16) holds, then the inequality (87) is fulfilled $\forall j \geq 0$. Firstly, since condition (16) means that $\chi < 1/2$ in (86), we have:

$$m\frac{\mu^{(j)}}{\gamma(\delta^{(j)})^{\nu+1}} = m\frac{\mu}{\gamma\delta^{\nu+1}}\chi^{2^j-1} \leq m\frac{\mu}{\gamma\delta^{\nu+1}}\left(\frac{1}{2}\right)^j \leq m\frac{\mu}{\gamma\delta^{\nu+1}} < \frac{1}{2}, \qquad \forall j \geq 0.$$

Moreover, we also have:

$$\sum_{j=0}^{+\infty} m\frac{\mu^{(j)}}{\gamma(\delta^{(j)})^{\nu+1}} \leq m\frac{\mu}{\gamma\delta^{\nu+1}}\sum_{j=0}^{+\infty}\left(\frac{1}{2}\right)^j \leq 2m\frac{\mu}{\gamma\delta^{\nu+1}} < \Theta,$$
$$\sum_{j=0}^{+\infty} m\frac{\mu^{(j)}}{\gamma(\delta^{(j)})^{\nu}} \leq 2m\frac{\mu}{\gamma\delta^{\nu}} < \Theta.$$

To verify inequalities of (87) involving the definition of $\Theta^{(j)}$, we proceed by induction with respect to n. The fist step is immediate, since condition (16) implies that (87) holds for $n = 1$ (i.e. for $j = 0$). Let us assume that (87) is satisfied up to some $n \geq 1$. This fact implies that we can iterate n times and that if we define $\Lambda^{(j)}$ as Λ in (84), but by replacing τ, $\tau^{(1)}$, N, $N^{(1)}$, S, $S^{(1)}$, and $\rho^{(1)}$ by $\tau^{(j)}$, $\tau^{(j+1)}$, $N^{(j)}$, $N^{(j+1)}$, $S^{(j)}$, $S^{(j+1)}$, and $\rho^{(j+1)}$, respectively, then we have [compare with (85)]:

$$\|\tau^{(j+1)} - \tau^{(j)}\|_{\rho^{(j+1)}} \leq m\frac{\mu^{(j)}}{\gamma(\delta^{(j)})^{\nu}}, \qquad \Lambda^{(j)} \leq m\frac{\mu^{(j)}}{\gamma(\delta^{(j)})^{\nu+1}}, \qquad j = 0,\ldots,n-1. \tag{88}$$

These bounds mean that, in particular, we can control the size of $\|D\tau^{(n)}\|_{\rho^{(n)}}$ by the following sum:

$$\|D\tau^{(n)}\|_{\rho^{(n)}} \leq \|D\tau^{(0)}\|_{\rho^{(0)}} + \sum_{j=0}^{n-1}\|D\tau^{(j+1)} - D\tau^{(j)}\|_{\rho^{(j+1)}} \leq \|D\tau\|_{\rho}$$
$$+ \sum_{j=0}^{n-1} m\frac{\mu^{(j)}}{\gamma(\delta^{(j)})^{\nu+1}}.$$

Consequently, by using the definition of Θ in (13), we obtain:

$$m\frac{\mu^{(n)}}{\gamma(\delta^{(n)})^{\nu+1}} \le \|D\tau\|_\rho - \|D\tau^{(n)}\|_{\rho^{(n)}} + \sum_{j=0}^{n} m\frac{\mu^{(j)}}{\gamma(\delta^{(j)})^{\nu+1}}\Theta$$
$$< \|D\tau\|_\rho - \|D\tau^{(n)}\|_{\rho^{(n)}} + \le m_3 - \|D\tau^{(n)}\|_{\rho^{(n)}}.$$

By performing analogous computations for $\|(N^{(n)})^{-1}\|_{\rho^{(n)}}$, $d_{\tau^{(n)},\rho^{(n)},\mathcal{U}}$, and $|(\langle S^{(n)}\rangle_\theta)^{-1}|$ we verify that $m\mu^{(n)}/(\gamma(\delta^{(n)})^{\nu+1}) < \Theta^{(n)}$. Only the control of $d_{\tau^{(n)},\rho^{(n)},\mathcal{U}}$ leads to a slightly different case with respect to $\|D\tau^{(n)}\|_{\rho^{(n)}}$, since it involves the bounds of $\|\tau^{(j+1)} - \tau^{(j)}\|_{\rho^{(j+1)}}$ in (88) instead of the bounds of $\Lambda^{(j)}$. Therefore, we make this case equivalent to the other ones by using that $1/(\delta^{(n)})^\nu \le 1/(\delta^{(n)})^{\nu+1}$. Consequently, the inequality (87) also holds for $j = n$, so we can iterate again. Hence, results on the statement of Theorem 2.6 follow by taking the limit $n \to +\infty$. In particular, we have:

$$\|\tau^* - \tau\|_{\rho^*} \le \sum_{j=0}^{+\infty} \|\tau^{(j+1)} - \tau^{(j)}\|_{\rho^{(j+1)}} \le \sum_{j=0}^{+\infty} m\frac{\mu^{(j)}}{\gamma(\delta^{(j)})^\nu} \le 2m\frac{\mu}{\gamma\delta^\nu},$$

and similarly for $\|D\tau^* - D\tau\|_{\rho^*}$.

Acknowledgements We thank the anonymous reviewers for their careful reading of the manuscript and valuable comments, which helped to improve the manuscript.

References

Arnold, V.I.: Proof of a theorem of A. N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations. Russ. Math. Surv. **18**(5), 9–36 (1963)

Benettin, G., Galgani, L., Giorgilli, A., Strelcyn, J.-M.: A proof of Kolmogorov's theorem on invariant tori using canonical transformations defined by the Lie method. Nuovo Cimento B (11) **79**(2), 201–223 (1984)

Broer, H.W., Sevryuk, M.B.: KAM theory: Quasi-periodicity in dynamical systems. In: Broer, H.W., Hasselblatt, B., Takens, F. (eds.) Handbook of Dynamical Systems, vol. 3C, pp. 249–344. Elsevier (2010)

Calleja, R., Celletti, A., de la Llave, R.: A KAM theory for conformally symplectic systems: efficient algorithms and their validation. J. Diff. Equ. **255**(5), 978–1049 (2013)

Calleja, R., de la Llave, R.: A numerically accessible criterion for the breakdown of quasi-periodic solutions and its rigorous justification. Nonlinearity **23**(9), 2029–2058 (2010)

Celletti, A., Chierchia, L.: On the stability of realistic three-body problems. Comm. Math. Phys. **186**(2), 413–449 (1997)

de la Llave, R.: A tutorial on KAM theory. In: Smooth Ergodic Theory and its Applications (Seattle, WA, 1999), volume 69 of Proc. Sympos. Pure Math., pages 175–292. Amer. Math. Soc., Providence (2001)

de la Llave, R., González, A., Jorba, À., Villanueva, J.: KAM theory without action-angle variables. Nonlinearity **18**(2), 855–895 (2005)

Fontich, E., de la Llave, R., Sire, Y.: Construction of invariant whiskered tori by a parameterization method. Part I: maps and flows in finite dimensions. J. Diff. Equ. **246**, 3136–3213 (2009)

Gabern, F., Jorba, À., Locatelli, U.: On the construction of the Kolmogorov normal form for the Trojan asteroids. Nonlinearity **18**(4), 1705–1734 (2005)

González, A., de la Llave, R.: Analytic smoothing of geometric maps with applications to KAM theory. J. Diff. Equ. **245**(5), 1243–1298 (2008)

González-Enríquez, A., Haro, A., de la Llave, R.: Singularity theory for non-twist KAM tori. Mem. Am. Math. Soc. **227**(1067), vi+115 (2014)

Haro, À., Canadell, M., Figueras, J.-L., Luque, A., Mondelo, J.M.: The Parameterization Method for Invariant Manifolds, Volume 195 of Applied Mathematical Sciences. Springer [Cham], New York City (2016). (From rigorous results to effective computations). http://link.springer.com/content/pdf/bfm%3A978-3-319-29662-3%2F1.pdf

Huguet, G., de la Llave, R., Sire, Y.: Computation of whiskered invariant tori and their associated manifolds: new fast algorithms. Discrete Contin. Dyn. Syst. **32**(4), 1309–1353 (2012)

Jorba, À., de la Llave, R., Zou, M.: Lindstedt series for lower-dimensional tori. In: Simó, C. (ed.) Hamiltonian Systems with Three or More Degrees of Freedom (S'Agaró, 1995), volume 533 of NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., pp. 151–167. Kluwer Acad. Publ., Dordrecht (1999)

Jorba, À., Villanueva, J.: On the normal behaviour of partially elliptic lower-dimensional tori of Hamiltonian systems. Nonlinearity **10**(4), 783–822 (1997)

Jorba, À., Villanueva, J.: On the persistence of lower-dimensional invariant tori under quasi-periodic perturbations. J. Nonlinear Sci. **7**(5), 427–473 (1997)

Jorba, À., Villanueva, J.: Numerical computation of normal forms around some periodic orbits of the restricted three-body problem. Phys. D **114**(3–4), 197–229 (1998)

Kolmogorov, A.N.: On conservation of conditionally periodic motions for a small change in Hamilton's function. Dokl. Akad. Nauk SSSR (N.S.), **98**, 527–530 (1954). Translated in pp. 51–56 of Stochastic Behavior in Classical and Quantum Hamiltonian Systems, Como 1977 (eds. G. Casati and J. Ford) Lect. Notes Phys. 93, Springer, Berlin (1979)

Locatelli, U., Giorgilli, A.: Invariant tori in the secular motions of the three-body planetary systems. Celestial Mech. Dynam. Astronom **78**(1–4), 47–74 (2001). (2000. New developments in the dynamics of planetary systems (Badhofgastein, 2000))

Luque, A., Villanueva, J.: A KAM theorem without action-angle variables for elliptic lower dimensional tori. Nonlinearity **24**(4), 1033–1080 (2011)

Luque, A., Villanueva, J.: A numerical method for computing initial conditions of Lagrangian invariant tori using the frequency map. Phys. D **325**, 63–73 (2016)

Moser, J.: On invariant curves of area-preserving mappings of an annulus. Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II **1962**, 1–20 (1962)

Moser, J.: A rapidly convergent iteration method and non-linear differential equations. II. Ann. Scuola Norm. Sup. Pisa (3) **20**, 499–535 (1966)

Moser, J.: A rapidly convergent iteration method and non-linear partial differential equations. I. Ann. Scuola Norm. Sup. Pisa (3) **20**, 265–315 (1966)

Neĭshtadt, A.I.: Estimates in the Kolmogorov theorem on conservation of conditionally periodic motions. J. Appl. Math. Mech. **45**(6), 1016–1025 (1981)

Perry, A.D., Wiggins, S.: KAM tori are very sticky: rigorous lower bounds on the time to move away from an invariant Lagrangian torus with linear flow. Phys. D **71**(1–2), 102–121 (1994)

Pöschel, J.: Integrability of Hamiltonian systems on Cantor sets. Comm. Pure Appl. Math. **35**(5), 653–696 (1982)

Pöschel, J.: A lecture on the classical KAM theorem. In: Smooth Ergodic Theory and its Applications (Seattle, WA, 1999), vol. 69 of Proc. Sympos. Pure Math., pp. 707–732. Amer. Math. Soc., Providence (2001)

Rüssmann, H.: On optimal estimates for the solutions of linear partial differential equations of first order with constant coefficients on the torus. In: Dynamical Systems, Theory and Applications (Rencontres, Battelle Res. Inst., Seattle, Wash., 1974), pp. 598–624. Lecture Notes in Phys., Vol. 38. Springer, Berlin (1975)

Rüssmann, H.: On a new proof of Moser's twist mapping theorem. In: Proceedings of the Fifth Conference on Mathematical Methods in Celestial Mechanics (Oberwolfach, 1975), Part I. Celestial Mech. **14**(1):19–31 (1976)

Salamon, D.: The Kolmogorov-Arnold-Moser theorem. Math. Phys. Electron. J., 10:Paper 3, 37 pp. (electronic) (2004)

Salamon, D., Zehnder, E.: KAM theory in configuration space. Comment. Math. Helv. **64**(1), 84–132 (1989)

Villanueva, J.: Kolmogorov theorem revisited. J. Diff. Equ. **244**(9), 2251–2276 (2008)

Weinstein, A.: Lectures on Symplectic Manifolds, Volume 29 of CBMS Regional Conference Series in Mathematics. American Mathematical Society, Providence (1979)

Zehnder, E.: Generalized implicit function theorems with applications to some small divisor problems. II. Comm. Pure Appl. Math. **29**(1), 49–111 (1976)