A parameterization method for Lagrangian tori of exact symplectic maps of $\mathbb{R}^{2r}$ *

Jordi Villanueva[†]

Departament de Matemàtiques,

Universitat Politècnica de Catalunya,

Diagonal 647, 08028 Barcelona (Spain).

February 1, 2018

Abstract

We are concerned with analytic exact symplectic maps of $\mathbb{R}^{2r}$ endowed with the standard symplectic form. We study the existence of a real analytic torus of dimension r, invariant by the map and carrying quasi-periodic motion with a prefixed Diophantine rotation vector. Therefore, this torus is a Lagrangian manifold. We address the problem by the parameterization method in KAM theory. The main aspect of our approach is that we do not look for the parameterization of the torus as a solution of the corresponding invariance equation. Instead, we consider a set of three equations that, all together, are equivalent to the invariance equation. These equations arise from the geometric and dynamical properties of the map and the torus. Suppose that an approximate solution of these equations is known and that a suitable non-degeneracy (twist) condition is satisfied. Then, this system of equations is solved by a quasi-Newton-like method, provided that the initial error is sufficiently small. By quasi-Newton-like we mean that the convergence is almost quadratic, but that at each iteration we have to solve a nonlinear equation. Although it is straightforward to build a quasi-Newton method for the selected set of equations, proceeding in this way we improve the convergence condition. The selected definition of error reflects the level at which the error associated with each of these three equations contributes to the total error. The map is not required to be close to integrable or to be expressed in action-angle variables. Suppose the map is ε-close to an integrable one and consider the portion of the phase space not filled up by Lagrangian invariant tori of the map. Then, the upper bound for the Lebesgue measure of this set that we may predict from the result is of $\mathcal{O}(\varepsilon^{1/2})$. In light of the classical KAM theory for exact symplectic maps, an upper bound of $\mathcal{O}(\varepsilon^{1/2})$ for this measure is the expected estimate. The result also has some implications for finitely differentiable maps.

MSC: 37J10; 37J40; 70K43.
Keywords: KAM theory; Exact symplectic maps; Lagrangian tori; Parameterization methods.

1   Introduction

The problem of the existence of invariant tori carrying quasi-periodic motion, in many contexts of the dynamical systems, constitutes the core of the celebrated KAM theory (named after the pioneering works of Kolmogorov, Arnol'd and Moser [1, 18, 21]). Within the framework of the classical KAM theory, this problem is usually approached by transformation theory applied to a system that is a small perturbation of one that has a specific

[†]E-mail: jordi.villanueva@upc.edu; web-page: https://www.mat.upc.edu/en/people/jordi.villanueva

structure. We mean that the unperturbed system is written in a coordinate system that contains as many angular variables as the dimension of the tori, and that the unperturbed tori are obtained by giving specific constant values to the other variables. Then, the persistence of these tori in the perturbed system is addressed by applying a sequence of transformations to the system. Typically, this sequence has an almost quadratic rate of convergence and, at the limit, we also obtain invariant tori of the complete system by giving specific constant values to some variables. Roughly speaking, by almost quadratic convergence speed we mean that it is quadratic modulo the effect of the small divisors associated to the frequencies and of the Cauchy estimates performed to control some derivatives (in the analytic context). We refer the interested reader to [2, 6] for some nice surveys of KAM theory.

Fundamental examples of the classical approach to the KAM theory are the computation of Lagrangian tori of a nearly-integrable Hamiltonian H_ε and of an exact symplectic map F_ε close to an integrable one. In both contexts, the coordinate frame $(\theta, I) \in \mathbb{T}^r \times \mathbb{R}^r$ is taken as a symplectic system of action-angle variables of the r-dimensional annulus[1]. Then, $H_\varepsilon(\theta, I) = h(I) + \varepsilon\, f(\theta, I)$ and $F_\varepsilon(\theta, I) = \mathcal{F}(\theta, I) + \varepsilon\, G(\theta, I)$ are ε-small perturbations of integrable systems. Specifically, we mean that $\mathcal{F}$ is a skew-product of $\mathbb{T}^r \times \mathbb{R}^r$ of the form:

$$\mathcal{F}(\theta, I) = (\theta + \nabla\psi(I), I), \tag{1}$$

where $\psi(I)$ is a scalar function and ∇ denotes the gradient operator.

The application to particular examples of KAM results based on transformation theory presents several drawbacks, both from the theoretical viewpoint and from the point of view of the applications (e.g., if we want to numerically implement the proof in a computer). A natural example is the computation of a Lagrangian torus, carrying quasi-periodic motion, of a given Hamiltonian H written in Cartesian coordinates. To address the problem by transformation theory, firstly we need to find an integrable approximation H_0 of H. In fact, it is sufficient to find a nearby Hamiltonian H_0 which has an invariant Lagrangian torus. We refer to the distance between H and H_0 as the size of the perturbation. Then, we have to express H in action-angle variables adapted to an invariant torus of H_0 and to apply to H a sequence of canonical transformations. However, even if we can express H in action-angle coordinates, generally we can only ensure convergence of the KAM normalizing procedure for extremely small values of the perturbation. Besides, implementing canonical transformations in a computer is quite expensive from the computational point of view (although performing canonical transformations on a Hamiltonian is perhaps the best scenarios within the framework of transformation theory).

Drawbacks above have led to the development of new tools for the computation of invariant tori that avoid the use of transformation theory. The basic idea of this alternative approach is to look for invariant tori through a parameterization. The target parameterization τ constitutes an embedding between $\mathbb{T}^r$ and an invariant torus $\mathcal{T}$, where $r = \dim \mathcal{T}$. Since we are concerned with invariant tori carrying quasi-periodic motion, it is natural to look for a parameterization so that the pull-back by τ of the dynamics on $\mathcal{T}$ is a parallel dynamics on $\mathbb{T}^r$. We refer to the equation for τ arising from these conditions as the invariance equation of the parameterization. This equation is then solved by a quasi-Newton method. By a quasi-Newton method for τ we mean that only the parameterization is modified iteratively, but not the system, and that the convergence rate is almost quadratic with the step. This approach is usually referred to as as the parameterization method in KAM theory.

Some crucial aspects of parameterization methods are summarized below. We do not need the system to be a perturbation of one with a specific structure or to be written in an adapted coordinate frame. Instead, what has to be known is the parameterization of a quasi-torus of the system, i.e., a parameterization that results in a sufficiently small error when the invariance equation is evaluated in it. This quasi-torus may come from a wide range of sources, that depend on the context at hand. E.g., from any numerical method for computing invariant tori in the literature (see [20] for an overview of methods). The proofs are constructive and take advantage of the geometric and dynamical properties of the systems to solve, modulo quadratic error terms, the linearization of the invariance equation around the parameterization of a given quasi-torus. Since proofs are constructive, parameterization

---

[1] As customary in the literature, we will use the definition $\mathbb{T}^r = (\mathbb{R}/2\pi\mathbb{Z})^r$ for flows and the definition $\mathbb{T}^r = (\mathbb{R}/\mathbb{Z})^r$ for maps.

methods can be implemented in a computer, leading to efficient algorithms for computing invariant tori. These algorithms are computationally less demanding than those based on transformation theory or on brute force methods for solving the invariance equation (e.g., by performing a discretization of it using Fourier polynomials, finite differences, splines, etc.). Parameterization methods are also useful in the framework of the computer assisted proofs (CAPs). If we perform a CAP to address the rigorous verification of the existence of an invariant torus of a particular system, parameterization methods lead to a better threshold for the size of the initial error that ensures the existence of the torus than methods based on transformation theory. Moreover, before performing the CAP, the numerical implementation of the proof can be used to previously improve the approximate parameterization. Finally, we point out that although analytic systems are the natural context for performing functional iterations in KAM theory, parameterization results can be extended to finitely differentiable systems if we combine the proofs for the analytic case with analytic smoothing techniques.

The foundational ideas of parameterization methods in KAM theory were introduced in [7]. This work discusses the existence of Lagrangian tori, carrying quasi-periodic motion, for both analytic Hamiltonian systems and for analytic exact symplectic maps (the discrete-time version of Hamilton's mechanics). The approach of [7] follows the steps marked by previous works developed by several authors (see [5, 17, 22, 23, 27, 29, 32]). As is implicit in parameterization methods, results in [7] do not require the system to be a perturbation of an integrable one or to be written in action-angle coordinates. Actually, the reference frame is a symplectic system of Cartesian coordinates. Later on, the approach of [7] has been successfully extended to the computation of invariant tori in several contexts of the dynamicals systems, as well as to address some aspects mentioned above in connection to parameterization methods (as numerical methods, CAPs and analytic smoothing). See [3, 4, 8, 9, 10, 11, 14, 15, 16, 19]. We refer the interested reader to the survey [13] for more details and references.

Let us consider a real analytic Hamiltonian h, with r degrees of freedom, and $\omega \in \mathbb{R}^r$. The results in [7] establish the existence of a Lagrangian invariant torus of h carrying quasi-periodic motion with frequency vector ω. Hypotheses are a Diophantine condition for ω, the knowledge of the parameterization τ of a quasi-torus of h (for which the invariance error e is sufficiently small), and that certain non-degeneracy (twist) condition is satisfied. The Diophantine conditions for the frequency vector ω of a vector-field are:

$$|\langle k, \omega \rangle| \geq \gamma |k|_1^{-\nu}, \qquad \forall k \in \mathbb{Z}^r \setminus \{0\}, \tag{2}$$

for some $\nu \geq r - 1$ and $\gamma > 0$, where $\langle \cdot, \cdot \rangle$ denotes the inner product and $|k|_1 = |k_1| + \cdots + |k_r|$. The invariance error of the parameterization τ is defined as $e = L_\omega \tau - J \nabla h(\tau)$, where L_ω is the Lie derivative with respect to ω and J is the matrix of the canonical 2-form of $\mathbb{R}^{2r}$ (see (5)). The twist condition, which allows us to preserve the value of ω iteratively, is defined in terms of the Hessian of h evaluated at τ. For the invariant torus $\mathbb{T}^r \times \{I_0\}$ of an integrable Hamiltonian $h(I)$, written in action-angle variables $(\theta, I) \in \mathbb{T}^r \times \mathbb{R}^r$, this twist condition reads as the Kolmogorov's non-degeneracy condition $\det(D^2 h(I_0)) \neq 0$. Hence, this result is formulated with the same aim as the classical Kolmogorov's theorem on the persistence of quasi-periodic motions [18].

The result of [7] for maps is completely analogous to the one for flows. The main ingredients are an exact symplectic map F of $\mathbb{R}^{2r}$, a rotation vector $\omega \in \mathbb{R}^r$, and the parameterization τ of a quasi-torus of F. The Diophantine conditions for ω are:

$$|\langle k, \omega \rangle - m| \geq \gamma |k|_1^{-\nu}, \qquad \forall k \in \mathbb{Z}^r \setminus \{0\}, \quad \forall m \in \mathbb{Z}, \tag{3}$$

for some $\nu \geq r$ and $\gamma > 0$. The twist condition for F reads as $\det(D^2 \psi(I_0)) \neq 0$ when formulated on the invariant torus $\mathbb{T}^r \times \{I_0\}$ of the integrable map (1) (see remark 2.8). The invariance error of τ is $e = F(\tau) - \tau \circ T_\omega$, where $T_\omega(\cdot) = \cdot + \omega$. A specific feature of the result of [7] for maps is that it is formulated for a general analytic exact symplectic structure of $\mathbb{R}^{2r}$, but not only for the canonical one.

We denote by $\Delta(\rho)$ the complex strip of width ρ around $\mathbb{T}^r$ and by $\| \cdot \|_\rho$ the sup norm in the set $\Delta(\rho)$ (see definition 2.4). Suppose that the parameterization τ of the initial quasi-torus can be analytically extended to the set

$\Delta(\rho)$, for some $\rho > 0$, and denote by $\varepsilon = \|e\|_\rho$ the size of the invariance error of τ in $\Delta(\rho)$. Roughly speaking, the convergence condition of [7] is satisfied, for both Hamiltonians and maps, if $\gamma^{-4}\rho^{-4\nu}\varepsilon$ is sufficiently small, where γ and ν are those of the Diophantine conditions of ω (see (2) and (3)). According to this result, we can ensure that the invariant torus exists provided that γ satisfies a lower bound of $\mathcal{O}(\sqrt[4]{\varepsilon})$. Indeed, when computing invariant tori of Hamiltonian systems and symplectic maps, in the framework of Kolmogorov's theorem, the usual convergence condition in the literature is a lower bound for γ of $\mathcal{O}(\sqrt[4]{\varepsilon})$. This assertion holds whether the KAM proof is performed using parameterization methods or transformation theory (in this latter case ε denotes the size of the perturbation). However, for Lagrangian tori of both Hamiltonians and symplectic maps, the natural result expected a priori is the convergence of the KAM iterations if a lower bound for γ of $\mathcal{O}(\sqrt{\varepsilon})$ is satisfied. Explicitly, by the celebrated KAM theorem [1], we know that if $H_\varepsilon(\theta, I) = h(I) + \varepsilon f(\theta, I)$ is a nearly-integrable Hamiltonian, then the Lebesgue measue of the set defined by the holes between invariant tori of H_ε has an upper bound of $\mathcal{O}(\sqrt{\varepsilon})$ (see [24, 25]). This is to say that the tori of the integrable system that survive to the perturbation are those with a Diophantine frequency vector (2) for which γ is greater than $\mathcal{O}(\sqrt{\varepsilon})$.

With the aim of obtaining a lower bound for γ of $\mathcal{O}(\sqrt{\varepsilon})$, in [31] we re-visited the result of [7] for Lagrangian tori of Hamiltonian systems. The main idea behind the new approach of [31] is our feeling that dealing with the invariance error e, as a single source of error for the parameterization, may not be a good idea if we pursue sharp estimates. This feeling is motivated by a previous work [30] in which we re-visited Kolmogorov's theorem [18] (using transformation theory). Kolmogorov's theorem is formulated for a Hamiltonian H, written in action-angle variables $(\theta, I) \in \mathbb{T}^r \times \mathbb{R}^r$, which is a small perturbation of a Hamiltonian h in Kolmogorov's normal form. This reduced form is the more general expression for h from which we can ensure that the Lagrangian torus $\mathbb{T}^r \times \{I_0\}$, for some $I_0 \in \mathbb{R}^r$, is invariant and carries parallel flow in θ (i.e., $\dot{\theta} = \omega$, for some frequency vector ω). To control the distance of H to Kolmogorov's normal form we have two natural sources of error. Explicitly, the (scalar) error in the energy $e_1(\theta) = H(\theta, 0) - \langle H(\theta, 0)\rangle_\theta$, where $\langle \cdot \rangle_\theta$ is the average with respect to $\theta \in \mathbb{T}^r$, and the (vector) error in the frequencies $e_2(\theta) = \nabla_I H(\theta, 0) - \omega$. The customary in the literature is to measure the distance of H to Kolmogorov's normal form by the maximum of both errors, i.e., the size of the perturbation is $\varepsilon = \max\{\|e_1\|_\rho, \|e_2\|_\rho\}$, for some $\rho > 0$. Kolmogorov introduced a normalizing process for H (perfomed by a sequence of canonical transformations) which leads to a limit Hamiltonian that is also in Kolmogorov's normal form (with the same ω). If ε is as above, the convergence conditions arising in the literature are of the form $\gamma^{-4}\rho^{-s}\varepsilon$ sufficiently small (the specific value of $s > 0$ depends on the authors), which means convergence if γ has a lower bound of $\mathcal{O}(\sqrt[4]{\varepsilon})$. By contrast, in [30] we dealt with the size of both errors e_1 and e_2 separately. In this way, we established a convergence condition of the form $\gamma^{-1}\rho^{-\nu-5}(\|e_2\|_\rho + \gamma^{-1}\rho^{-\nu}\|\nabla e_1\|_\rho)$ sufficiently small. In particular, if ε is as above, we conclude that the invariant torus $\mathbb{T}^r \times \{I_0\}$ of h persists if γ satisfies a lower bound of $\mathcal{O}(\sqrt{\varepsilon})$. The main aspects that lead to improve the lower bound for γ are summarized as follows. First, in [30] the distance of H to Kolmogorov's normal form is controlled in terms of an appropriate weighted error. The definition of this error takes into account that the cohomological equations of Kolmogorov's normalizing process have triangular structure. Second, the sequences of canonical transformations and transformed Hamiltonians are written using explicit formulas, as compact as possible. Finally, after each iteration, the expression of the cohomological equations of the KAM process is used to rewrite some components of the new errors in terms of the errors of the previous step, rather than in terms of the solutions of these equations.

In [31] we adapt the ideas of [30] to compute, by the parameterization method, Lagrangian tori of a Hamiltonian system h written in Cartesian coordinates. Explicitly, given a particular frequency vector ω, in [31] we do not control the error of the parameterization τ of a quasi-torus $\mathcal{T}$ by the invariance error e, but by three errors $\{e_j\}_{j=1}^3$. If these errors cancel out simultaneously, then $\mathcal{T}$ is invariant. The scalar function e_1 indicates how far $\mathcal{T}$ is from belonging to an energy level set of h, the vector function e_2 measures the distance of $\mathcal{T}$ from being a Lagrangian manifold, and e_3 is a vector function that controls the error in the inner dynamics of $\mathcal{T}$. We note that, while in [30] we considered two errors, in [31] we use three. This is because using action-angle coordinates and transformation theory (as done in [30]), all the tori of the sequence of quasi-torus that we generate by the Kolmogorov's

normalizing procedure are Lagrangian manifolds in a straightforward way. If we deal with Cartesian coordinates, and we perform a parameterization method, it is unclear how to generate a sequence of parameterizations so that the corresponding quasi-tori are Lagrangian manifolds. This is the reason for which we should consider the error function e_2. In [31] we introduced a modified version of the parameterization method for which the convergence condition reads as $\gamma^{-1}\rho^{-\nu-1}(\|e_3\|_\rho + \gamma^{-1}\rho^{-\nu}(\|\nabla e_1\|_\rho + \|e_2\|_\rho))$ sufficiently small (ρ is the width of analyticity of the parameterization of the initial quasi-torus). Since in any perturbative scenario (i.e., if h is ε-close to a system that has an invariant torus) we expect the size of each e_j to be of order ε, the convergece condition holds if a lower bound for γ of $\mathcal{O}(\sqrt{\varepsilon})$ is satisfied. A natural context for this comment is when h is nearly-integrable. To apply the result of [31] to h we have to parameterize the tori of the integrable approximation, but not to introduce an adapted system of action-angle variables. Besides dealing with suitable projections of the invariance error, the use of expressions as compact as possible of the parameterizations plays a very important role in [31]. By compactness we mean replacing, where it is possible, addition of functions by composition of functions (see definition 3.6 for details). As a summary, to perform an iteration of the method of [31], the cohomological equations to be solved are analogous to those of [7], but the sequence of parameterizations is not the same.

In this work we extend the modified approach to the parameterization method of [31] to maps. We consider an exact symplectic real analytic map F of $\mathbb{R}^{2r}$, written in Cartesian coordinates and endowed with the standard symplectic form, and a vector $\omega \in \mathbb{R}^r$ that verifies the Diopantine conditions (3). We do not suppose that F is close to an integrable map or that we know a nearby map that has an invariant torus. In theorem 2.6 we compute the parameterization of a Lagrangian torus, invariant by F and carrying quasi-periodic motion with rotation vector ω. As done in [31], to formulate this result we do not control the error of the parameterization τ of a quasi-torus $\mathcal{T}$ of F by the invariance error e, but by a set of three errors $\{e_j\}_{j=1}^3$ (one scalar function and two vector functions). The equations for τ that define these error functions are discussed in proposition 3.2. We refer the reader to the comments below the proof of proposition 3.2 for the explanation of the geometric and dynamical meaning of these equations. The result of proposition 3.2 is that if $e_j = 0$, $\forall j = 1, 2, 3$, and the size of the Jacobian matrix of e is sufficiently small, then $\mathcal{T}$ is a Lagrangian torus, invariant by F, and carries quasi-periodic motion with rotation vector ω. Therefore, to apply proposition 3.2, not only the errors $\{e_j\}_{j=1}^3$ should be zero. We also have to check an additional condition for the norm of De. This condition does not ask this norm to be zero, but reasonably small. Although we cannot speak of a geometric constraint (since it involves the invariance error e), the Diophantine properties of the rotation vector ω do not play any role in this additional condition (to apply proposition 3.2 we only need ω to be non-resonant). We do not know how to get rid of this extra condition (at least not in a simple way). However, if we look at the statement of theorem 2.6, we realize that the resulting condition for the norm of De (see (15)) is very mild when compared with the smallness conditions we have to ask for $\{e_j\}_{j=1}^3$ (see (20)). Indeed, if we consider a quasi-torus of F for which this mild condition for De is satisfied, then we can use this property to relate the size of e to the size of $\{e_j\}_{j=1}^3$ (see equations from (77) to (87) in the proof of theorem 2.6). Consequently, if this mild condition for the norm of De is satisfied and the errors $\{e_j\}_{j=1}^3$ are small, a posteriori we can establish a bound for the norm of De which is much more small that the supposed a priori. In addition to small errors, hypotheses of theorem 2.6 also involve the same non-degeneracy (twist) condition introduced in [7].

The iterative method for τ, that constitutes the core of the modified approach to the parameterization method for F, is introduced in definition 3.6. As done in [31], we improve the compactness of the expression of the new approximation to the parameterization by using, where it is possible, composition of functions rather than addition of functions. The main difference with respect to the Hamiltonian context is that one of the small-divisors equations to be solved to perform definition 3.6 is a non-linear equation. Hence, we cannot refer to the equations in (41) as the cohomological equations of the method. Therefore, we use the terminology of generalized cohomological equations to refer to them. Besides, neither it it is correct to speak of a quasi-Newton method when referring to the construction of definition 3.6. The terminology of a quasi-Newton-like method seems most appropriate.

Formulating a nonlinear equation for b in (41) is the way we have found to ensure that the iterative errors behave appropriately. Indeed, the equations for a and b that follow naturally from the selected set of equations

for τ are the (linear) cohomological equations displayed in equations (44) and (45). However, suppose we apply the method of definition 3.6, but by replacing the equations in (41) by these linear equations. Then, after each iteration, the expressions of the new errors contain some terms whose size is controlled by the product of the sizes of a and b. As a consequence, the best convergence condition we have been able to establish for the iteration of this simplified construction is similar to that obtained in [7]. Although to perform the method we consider a nonlinear equation for b, it can be solved iteratively in a fairly simple way. The specific quasi-Newton method we use to compute b (and $\langle a \rangle_\theta$) is outlined at the bottom of definition 3.6 and the convergence details are discussed in proposition 3.7. The convergence condition of proposition 3.7 is consistent with the rest of conditions necessary to iterate the method of definition 3.6. Moreover, the estimates for b provided by proposition 3.7 are of the same order of magnitude to those we would obtain for b as a solution of the simplified (linear) equation (45).

To use theorem 2.6 from the theoretical viewpoint (e.g., to perform a CAP), the fact that the proof involves solving a nonlinear equation at each iteration does not really bother us. To apply definition 3.6 numerically, perhaps it may be enough to approximately solve the equation for b by performing (at least) one iteration of the method at the bottom of definition 3.6. For small errors, it is reasonable to expect that just a single iteration may provide an expression for b with numerical features similar to computing b as the solution of the nonlinear equation in (41).

In terms of the errors associated to the initial quasi-torus, the convergence condition of theorem 2.6 is, roughly speaking, $\gamma^{-1}\rho^{-\nu-2}(\|e_3\|_\rho + \gamma^{-1}\rho^{-\nu}(\|\nabla e_1\|_\rho + \|e_2\|_\rho))$ sufficiently small. If we are in a perturbative scenario, in which all these errors are of $\mathcal{O}(\varepsilon)$, then this condition means convergence provided that γ fulfills a lower bound of $\mathcal{O}(\sqrt{\varepsilon})$. Besides the significance that the exponent of γ has on this lower bound, the exponent of ρ of this condition has some interesting implications on the regularity of the tori for finitely differentiable maps (see remark 2.9).

The contents of the paper are organized as follows. In section 2, we present theorem 2.6, that constitutes the main result of the paper. Our modified approach to the parameterization method is introduced in section 3. Explicitly, in proposition 3.2 we construct the set of three equations we use to characterize a Lagrangian invariant torus of an exact symplectic map, in definition 3.6 we formulate the quasi-Newton-like method to solve this system of equations, and in proposition 3.7 we prove that there is a solution of the the generalized cohomological equations of definition 3.6. Section 4 is devoted to the proof of theorem 2.6.

2   Formulation of the main result

We consider the Euclidean space $\mathbb{R}^{2r}$ endowed with the symplectic structure defined by the canonical form $dx \wedge dy$, where $z = (x, y) \in \mathbb{R}^{2r}$ are Cartesian coordinates. Let $F : U \subset \mathbb{R}^{2r} \to \mathbb{R}^{2r}$ be a real analytic map, defined in an open set U. Suppose F is exact symplectic, i.e., that there is an analytic function $W : U \subset \mathbb{R}^{2r} \to \mathbb{R}$ (usually referred to as the primitive function of F) such that $F^* y dx = y dx + dW$. Apart from being analytic and exact symplectic, we do not suppose that F takes any specific form. We also consider a vector $\omega \in \mathbb{R}^r$, which satisfies the Diophantine conditions (3) for some $\gamma > 0$ and $\nu \geq r$. Our aim is to study the existence of a real analytic torus $\mathcal{T} \subset \mathbb{R}^{2r}$, of dimension r, invariant by F, and carrying quasi-periodic dynamics with rotation vector ω. Therefore, $\mathcal{T}$ should be a Lagrangian manifold (see defintion 2.3). To achieve this outcome, we rely on the parameterization method in KAM theory. Specifically, we compute a parameterization τ of $\mathcal{T}$ such that the the pull-back of the dynamics of F by τ is the translational map $T_\omega(\cdot) = (\cdot) + \omega$.

Definition 2.1. *For convenience, we will address the symplectic structure $dx \wedge dy$ in terms of the one-form $\varpi = y dx - x dy$ instead of $y dx$. Then, the map F is exact symplectic if $F^* \varpi = \varpi + dV$, where V is given by:*

$$V = 2W - F^* \langle x, y \rangle + \langle x, y \rangle = 2W - \langle F_x, F_y \rangle + \langle x, y \rangle,$$

where $F = (F_x, F_y)$. The coordinate expression of $F^* \varpi = \varpi + dV$ is given by the following formula:

$$(DF(z))^\top J F(z) = J z + \nabla V(z), \tag{4}$$

where $D(\cdot)$ stands for the Jacobian matrix of $(\cdot)$, ∇ is the gradient operator, $(\cdot)^\top$ denotes the transposition of $(\cdot)$, and J is the matrix representation of $dx \wedge dy$:

$$J = \begin{pmatrix} 0 & \mathrm{Id}_r \\ -\mathrm{Id}_r & 0 \end{pmatrix}. \tag{5}$$

We observe that J verifies $J^\top = -J$ and $J^2 = -\mathrm{Id}_{2r} = -\mathrm{Id}$. Since F is a symplectic map, the matrix DF also verifies $(DF)^\top J DF = J$.

Definition 2.2. *Let $\mathcal{T} \subset \mathbb{R}^{2r}$ be a real analytic torus of dimension r and $\tau : \mathbb{T}^r \to \mathbb{R}^{2r}$ an analytic parameterization of $\mathcal{T}$, where $\mathbb{T}^r = (\mathbb{R}/\mathbb{Z})^r$. Therefore, the transformation τ is an embedding between $\mathbb{T}^r$ and $\mathcal{T}$. We introduce the following definitions:*

$$N(\theta) = (D\tau(\theta))^\top D\tau(\theta), \quad \Omega(\theta) = (D\tau(\theta))^\top J D\tau(\theta), \quad \vartheta(\theta) = (D\tau(\theta))^\top J \tau(\theta), \quad \forall \theta \in \mathbb{T}^r. \tag{6}$$

Hence, N and Ω are the matrix representation of the pull-backs $\tau^*(\langle \cdot, \cdot \rangle)$ and $\tau^*(dx \wedge dy)$, respectively. The vector representation of $\tau^* \varpi$ is given by ϑ. We have that $N^\top = N$, $\Omega^\top = -\Omega$, and that $\det(N) \neq 0$ along the torus (since τ is an embedding). The relation $d\varpi = -2\, dx \wedge dy$ means that $2\,\Omega = D\vartheta - (D\vartheta)^\top$. This last expression implies, in particular, that $\langle \Omega \rangle_\theta = 0$, where $\langle (\cdot) \rangle_\theta = \int_{\mathbb{T}^r} (\cdot)\, d\theta$ is the average of any function defined in $\mathbb{T}^r$. If F is the exact symplectic map of definition 2.1, then equation (4) implies the following relation:

$$\vartheta(\theta) + \nabla(V(\tau(\theta))) = (D\tau(\theta))^\top (DF(\tau(\theta)))^\top JF(\tau(\theta)), \qquad \forall \theta \in \mathbb{T}^r. \tag{7}$$

We note in advance that if we wish to apply the approach of the paper to the parameterization τ of a quasi-torus $\mathcal{T}$ of F, then we have to pay attention to the expression of $\nabla(V(\tau(\theta)))$, but not to the expressions $V(\tau(\theta))$ and $\nabla V(\tau(\theta))$. Since we can compute $\nabla(V(\tau(\theta)))$ through the formula in (7), then the explicit expression of the primitive function V (or of W) is not involved in the statement of theorem 2.6 (although we can reformulate theorem 2.6 in terms of V, as noted in remark 2.7).

Definition 2.3. *We say that the torus $\mathcal{T}$ of definition 2.2 is invariant by the map F, and carries quasi-periodic dynamics with rotation vector ω, if there is a parameterization τ of $\mathcal{T}$ that verifies the following invariance equation:*

$$F(\tau(\theta)) = \tau(\theta + \omega), \qquad \forall \theta \in \mathbb{T}^r. \tag{8}$$

By taking partial derivatives of (8), we have that $DF(\tau(\theta))\, D\tau(\theta) = D\tau(\theta + \omega)$. Consequently, if τ is a solution of (8) and F is a symplectic map, then it is straightforward to verify that $\mathcal{L}_\omega \Omega = 0$, where we introduce the linear difference operator $\mathcal{L}_\omega f = f \circ T_\omega - f$. Conditions $\mathcal{L}_\omega \Omega = 0$ and $\langle \Omega \rangle_\theta = 0$ mean that $\Omega = 0$ (see (10)). Therefore, $\mathcal{T}$ is a Lagrangian manifold of $\mathbb{R}^{2r}$ and the columns of the matrices $JD\tau N^{-1}$ and $D\tau$ form a symplectic basis of $\mathbb{R}^{2r}$, at every point of the torus.

A crucial aspect of the computation of quasi-periodic Lagrangian tori of maps is the discussion of the following linear difference equation, where f and g may be scalar valued, vector valued or matrix valued:

$$\mathcal{L}_\omega f(\theta) = g(\theta), \qquad \forall \theta \in \mathbb{T}^r, \tag{9}$$

Given a function g, to solve equation (9) for f we expand g in Fourier series,

$$g(\theta) = \sum_{k \in \mathbb{Z}^r} \hat{g}(k)\, \mathrm{e}^{2\pi \mathrm{i} \langle k, \theta \rangle},$$

where $\hat{g}(0) = \langle g \rangle_\theta$. For further uses, we introduce the notation $\widetilde{g} = g - \hat{g}(0)$. If $\langle g \rangle_\theta = 0$ and ω verifies the non-resonance conditions $\langle k, \omega \rangle \neq m$, $\forall k \in \mathbb{Z}^r \setminus \{0\}$ and $\forall m \in \mathbb{Z}$, then equation (9) has infinitely many formal solutions for f. Explicitly, $f = \hat{f}(0) + \mathcal{L}_\omega^{-1} g$, where $\hat{f}(0)$ is free to take any value and

$$\mathcal{L}_\omega^{-1} g(\theta) = \sum_{k \in \mathbb{Z}^r \setminus \{0\}} \frac{\hat{g}(k)}{\mathrm{e}^{2\pi \mathrm{i}\langle k, \omega \rangle} - 1} \, \mathrm{e}^{2\pi \mathrm{i}\langle k, \theta \rangle}. \tag{10}$$

If $g = 0$, then f should be a constant function. The convergence of $\mathcal{L}_\omega^{-1} g$ for a generic g is strongly related to the regularity of g and to the lower bounds on the modulus of the "small divisors" $\mathrm{e}^{2\pi \mathrm{i}\langle k, \omega \rangle} - 1$. Since we are concerned with analytic functions and Diophantine vectors, we control $\mathcal{L}_\omega^{-1} g$ through the classical Rüssmann's estimates introduced in proposition 2.5 below. Firstly, we present the functional norms that we use in the paper.

Definition 2.4. *We denote by $|\cdot|$ the sup-norm of complex vectors and we extend the same notation to the compatible matrix norm. If $f : \mathcal{U} \subset \mathbb{C}^n \to \mathbb{C}$ is an analytic function, defined in an open set $\mathcal{U}$ and bounded in the closure of $\mathcal{U}$, then we introduce the norm $\|f\|_{\mathcal{U}} = \sup_{z \in \mathcal{U}} |f(z)|$. If f is vector valued (resp., matrix valued), then $\|f\|_{\mathcal{U}}$ is defined by computing the $|\cdot|$-norm of the constant vector (resp., matrix) defined by the $\|\cdot\|_{\mathcal{U}}$-norms of the entries of f. If f is a function defined in $\mathbb{T}^r$ (i.e., $f(\theta)$ is defined $\forall \theta \in \mathbb{R}^r$ and is 1-periodic in all the variables) that can be analytically extended to the complex strip $\Delta(\rho) = \{\theta \in \mathbb{C}^r : |\mathrm{Im}(\theta)| < \rho\}$, for some $\rho > 0$, then we introduce the specific notation $\|f\|_\rho = \|f\|_{\Delta(\rho)}$.*

Proposition 2.5 (Rüssmann's estimates [26, 28]). *Let g be a complex valued function (resp., vector valued, matrix valued) defined in $\mathbb{T}^r$, that can be analytically extended to $\Delta(\rho)$, for some $\rho > 0$, and that is bounded in the closure of $\Delta(\rho)$. Let $\omega \in \mathbb{R}^r$ be a Diophantine vector, which verifies (3) for some $\gamma > 0$ and $\nu \geq r$. If $\langle g \rangle_\theta = 0$, then equation (9) has a unique solution $f = \mathcal{L}_\omega^{-1} g$, which is analytic in $\Delta(\rho)$, is 1-periodic in all the variables and has zero average on $\mathbb{T}^r$. This solution is given by (10) and verifies the following estimates:*

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

where $\sigma \geq 1$ depends only on r and ν. Moreover, if g is real analytic, then so is f.

Theorem 2.6 below constitutes the main result of this work. See section 4 for the proof. To state it, we use notations introduced in definitions 2.1, 2.2, 2.3, and 2.4. The crucial point of theorem 2.6 is that if τ verifies the hypotheses of the statement, but the conditions $\nabla e_1 = e_2 = e_3 = 0$ are met, then the invariance error e cancels out (see proposition 3.2 and observe that $m_8 < 2$ implies that $\|(De)^\top D\tau_+ N_+^{-1}\|_\rho < 2$).

Theorem 2.6. *Let $F : U \subset \mathbb{R}^{2r} \to \mathbb{R}^{2r}$ be a map, defined in an open set U, and $\tau : \mathbb{T}^r \to \mathbb{R}^{2r}$ be a parameterization of a torus $\mathcal{T}$ of dimension r, such that $\mathcal{T} \subset U$. Suppose that F is exact symplectic with respect to $dx \wedge dy$ and that can be analytically extended to the open set $\mathcal{U} \subset \mathbb{C}^{2r}$, with $U \subset \mathcal{U}$. We also suppose that $\tau = \tau(\theta)$ can be analytically extended to the complex strip $\Delta(\rho)$, for some $\rho > 0$, with $\tau(\Delta(\rho)) \subset \mathcal{U}$. Let $\omega \in \mathbb{R}^r$ be a Diophantine vector, which verifies (3) for some $0 < \gamma \leq 1$ and $\nu \geq r$, and we denote by $\sigma = \sigma(r, \nu)$ the constant provided by proposition 2.5. Suppose that that the following estimates hold, for some constants $\{m_j\}_{j=1}^9$:*

$$\begin{aligned}
\|F\|_{\mathcal{U}} &\leq m_1, & \|DF\|_{\mathcal{U}} &\leq m_2, & \|D^2 F_j\|_{\mathcal{U}} &\leq m_3, && \forall j = 1, \ldots, 2r, & (11) \\
\|\tau\|_\rho &< m_4, & \|D\tau\|_\rho &< m_5, & \|N^{-1}\|_\rho &< m_6, & d_{\tau, \rho, \mathcal{U}} &> m_7 > 0, & (12)
\end{aligned}$$

where $F = (F_1, \ldots, F_{2r})$, $D^2(\cdot)$ is the Hessian matrix of the scalar function $(\cdot)$ and we introduce the notation

$$d_{\tau, \rho, \mathcal{U}} = \mathrm{dist}(\tau(\Delta(\rho)), \partial \mathcal{U}), \tag{13}$$

where $\partial \mathcal{U}$ stands for the boundary of $\mathcal{U}$ and $\mathrm{dist}(\cdot, \cdot)$ denotes the distance between two sets of $\mathbb{C}^{2r}$ (using the sup-norm). From F, τ, and ω, we introduce e and S, defined as:

$$e(\theta) = F(\tau(\theta)) - \tau(\theta + \omega), \qquad S(\theta) = (N(\theta + \omega))^{-1} (D\tau(\theta + \omega))^\top DF(\tau(\theta)) \, J \, D\tau(\theta) \, (N(\theta))^{-1}. \quad (14)$$

The vector function e is the error function associated to the invariance equation (8) and S is a r-dimensional square matrix function. We suppose that $\det(\langle S \rangle_\theta) \neq 0$ and that

$$\|(De)^\top\|_\rho \cdot m_5 \cdot m_6 \leq m_8 < 2, \qquad |(\langle S \rangle_\theta)^{-1}| < m_9. \quad (15)$$

From these quantities, we introduce $\Theta > 0$ defined as:

$$\Theta = \min \left\{ m_4 - \|\tau\|_\rho, m_5 - \|D\tau\|_\rho, m_6 - \|N^{-1}\|_\rho, d_{\tau, \rho, \mathcal{U}} - m_7, m_9 - |(\langle S \rangle_\theta)^{-1}| \right\}. \quad (16)$$

We compute the (error) functions e_1, e_2, and e_3, defined as

$$e_1(\theta) = (\tau(\theta + \omega))^\top J \, F(\tau(\theta)), \qquad e_3(\theta) = (D\tau(\theta + \omega))^\top (F(\tau(\theta)) - \tau(\theta + \omega)), \quad (17)$$

$$e_2(\theta) = (D\tau(\theta + \omega))^\top J \, \tau(\theta + \omega) - (D\tau(\theta))^\top (DF(\tau(\theta)))^\top J \, F(\tau(\theta)), \quad (18)$$

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

$$\|\nabla e_1\|_\rho \leq \mu_1, \qquad \|e_j\|_\rho \leq \mu_j, \quad j = 2, 3. \quad (19)$$

There is a constant $m \geq 1$, that depends only on r, ν, σ, and $\{m_j\}_{j=1}^9$, for which the following result holds. Given $0 < \delta < \min\{1, \rho/16\}$, we suppose that the quantities $\{\mu_j\}_{j=1}^3$ are small enough so that:

$$\frac{m}{\gamma \delta^{\nu+2}} \left(\mu_3 + \frac{\mu_1 + \mu_2}{\gamma \delta^\nu} \right) < \min \left\{ \frac{1}{2^{2(\nu+2)}}, \frac{m_8}{m_5 \cdot m_6}, \frac{\Theta}{2} \right\}. \quad (20)$$

Then, there is $\tau^* : \mathbb{T}^r \to \mathbb{R}^{2r}$ giving the analytic parameterization of a Lagrangian torus $\mathcal{T}^*$, which is invariant by F and carries quasi-periodic dynamics with rotation vector ω. Specifically, we have that $\mathcal{T}^* \subset U$, that τ^* can be analytically extended to the complex strip $\Delta(\rho^*)$, where $\rho^* = \rho - 16\delta$, with $\tau^*(\Delta(\rho^*)) \subset \mathcal{U}$, and that verifies:

$$F(\tau^*(\theta)) = \tau^*(\theta + \omega), \qquad \forall \theta \in \mathbb{T}^r,$$

$$\|\tau^* - \tau\|_{\rho^*} \leq \frac{2m}{\gamma \delta^\nu} \left(\mu_3 + \frac{\mu_1 + \mu_2}{\gamma \delta^\nu} \right), \qquad \|D\tau^* - D\tau\|_{\rho^*} \leq \frac{2m}{\gamma \delta^{\nu+1}} \left(\mu_3 + \frac{\mu_1 + \mu_2}{\gamma \delta^\nu} \right).$$

Furthermore, if N^* and S^* are the expressions defined by replacing τ by τ^* in the definitions of N and S, then:

$$\|\tau^*\|_{\rho^*} \leq m_4, \quad \|D\tau^*\|_{\rho^*} \leq m_5, \quad \|(N^*)^{-1}\|_{\rho^*} \leq m_6, \quad d_{\tau^*, \rho^*, \mathcal{U}} \geq m_7, \quad |(\langle S^* \rangle_\theta)^{-1}| \leq m_9.$$

Remark 2.7. *The proof of theorem 2.6 is constructive and can be implemented in a computer (see definition 3.6 for the quasi-Newton-like method used to compute τ^*, as well as the comments in the introduction regarding the approximate resolution of the nonlinear equation for b). In particular, in the proof we do not apply any change of coordinates to F. We do not give any explicit value for the constant m of the statement. However, since in the proofs of proposition 3.7 and theorem 2.6 we provide explicit formulas for all the involved expressions, it should not be difficult to generate a more explicit convergence condition. This makes the result also useful for performing CAPs. Since the exponents of γ and δ of the convergence condition are smaller than those of previous works, we hope that by using this result it may be possible to improve the treshold over the size of the errors ensuring convergence. We also oberve that if we know the explicit expression of the primitive function V of F (see definition 2.1), then can reformulate part of theorem 2.6 in terms of V and its derivatives. More specifically, we can use that $e_2 = \mathcal{L}_\omega \vartheta - \nabla(V(\tau))$ (see remark 3.4). We leave the rest of the details to the interested reader.*

Remark 2.8. *The matrix S of definition* (14), *usually referred to as the torsion matrix, and the non-degeneracy condition* $\det(\langle S \rangle_\theta) \neq 0$, *were introduced in [7]. For the invariant torus* $\mathcal{T} = \mathbb{T}^r \times \{I_0\}$ *of the integrable skew-product* $\mathcal{F}$ *of equation* (1), *parameterized by* $\tau(\theta) = (\theta, I_0)$, *this matrix is* $S = -D^2\psi(I_0)$. *Consequently, this non-degeneracy condition reads as the twist condition* $\det(D^2\psi(I_0)) \neq 0$, *which is the analogous of Kolmogorov's non-degeneracy condition for* $\mathcal{F}$ *at* $\mathcal{T}$. *The matrix* S *appears naturally when discussing the so-called automatic reducibility of quasi-periodic Lagrangian tori of an exact symplectic map (also introduced in [7]). For more details, we refer to equation* (27) *and related comments.*

Remark 2.9. *Theorem 2.6 is formulated for a map written in Cartesian coordinates, but its extension to a reference frame containing some angular variables is straightforward. As a matter of fact, if F is written in action-angle coordinates,* $(x, y) \in \mathbb{T}^r \times \mathbb{R}^r$, *and we look for Lagrangian tori homotopic to* $\mathbb{T}^r \times \{0\}$, *then the only modification worth to be mentioned is that the parameterizations should be of the form* $\tau(\theta) = (\theta + \bar{\tau}_x(\theta), \bar{\tau}_y(\theta)))$, *where* $\bar{\tau} : \mathbb{T}^r \to \mathbb{R}^{2r}$. *Beyond this simple observation, there are other generalizations and extensions of theorem 2.6 that we are convinced that can be also addressed successfully. Among them, we would like to mention the case in which the symplectic form is not the canonical one, but depends on the point (this is, indeed, the context considered in [7]). It is also possible to combine theorem 2.6 with analytic smoothing techniques to extend the result to finitely differentiable maps. This approach was already addressed in [10] regarding to the result of [7]. Specifically, the regularity estimates for the tori are strongly related to the exponent with which δ appears in the denominator of the convergence condition. We are also quite confident that the approach presented can be easily combined with the parametric KAM theory. Therefore, according to the selected equations for the torus (see proposition 3.2 and comments below the proof of the proposition), we should be able to deal with non-degeneracy conditions more general than the Kolmogorov's twist condition, as well as with the existence of Lagrangian tori of (non-exact) symplectic maps (by adapting the ideas of the translated torus theorem). For details on both contexts, we refer the interested reader to the monograph [11] and references therein.*

3   A new approach to the parameterization method

In this section we introduce the formal aspects of our approach to the parameterization method. Throughout the section we use definitions and notations introduced in section 2 (sometimes without explicit mention). We consider a real analytic exact symplectic map F of $\mathbb{R}^{2r}$ and a Diophantine rotation vector $\omega \in \mathbb{R}^r$. Given an analytic parameterization τ of a torus $\mathcal{T} \subset \mathbb{R}^{2r}$, of dimension r, we denote by e the invariance error of τ (see (14)):

$$e(\theta) = F(\tau(\theta)) - \tau(\theta + \omega). \tag{21}$$

Our goal is to find a parameterization τ^* for which the corresponding invariance error e^* is zero.

Remark 3.1. *In most of the formulas we omit the explicit dependence on θ of τ as well as of related expressions. Since ω is fixed, we introduce the notation f_+ to denote the function $f_+ = f \circ T_\omega$ (e.g., we write $e = F(\tau) - \tau_+$).*

Suppose we know the parameterization of a quasi-torus $\mathcal{T}$ of F, i.e., a parameterization τ for which $\|e\|_\rho \ll 1$, for some $\rho > 0$. Our aim is to correct τ iteratively in such a way the sequence of parameterizations thus obtained converges to τ^*. The natural approach to improve τ is to build a new parameterization $\tau^{(1)}$ through a quasi-Newton method. Therefore, we compute the linear approximation around τ of the invariance error of $\tau^{(1)}$, and we try to cancel out the resulting expression. If $\tau^{(1)} = \tau + \Delta\tau$, then we obtain the following (linear) equation for $\Delta\tau$:

$$\mathcal{R}(\Delta\tau) = -e, \tag{22}$$

where $\mathcal{R}(\cdot) = DF(\tau)(\cdot) - (\cdot)_+$ is a linear difference operator that can be applied to vector valued functions and matrix valued functions defined in $\mathbb{T}^r$ (with $2r$ rows). In [7], equation (22) is solved approximately (modulo

quadratic error terms on e and De). This solution for $\Delta\tau$ gives rise to a quasi-Newton method for equation (8). The crucial point of the construction of [7] is that if e is small then so is the matrix Ω (see definition 2.2). More concretely, since $\langle\Omega\rangle_\theta = 0$, the size of Ω can be controlled in terms of the size of De by the following expression:

$$\Omega = \mathcal{L}_\omega^{-1}(-(D\tau_+)^\top JDe + (De)^\top J\,DF(\tau)\,D\tau). \tag{23}$$

Formula (23) is immediate from the following relation

$$DF(\tau)\,D\tau = D\tau_+ + De \tag{24}$$

and the symplectic character of F. Since Ω is small, then the quasi-torus $\mathcal{T}$ is an approximate Lagrangian manifold and the columns of the matrices $JD\tau N^{-1}$ and $D\tau$ form an approximate symplectic basis of $\mathbb{R}^{2r}$, at every point of $\mathcal{T}$. Hence, we can express the (small) correction $\Delta\tau$ as:

$$\Delta\tau = JD\tau N^{-1}\,a + D\tau\,b, \tag{25}$$

for some r-dimensional vector functions a and b (to be determined). The smallness of Ω also implies that the action of the operator $\mathcal{R}$ on this approximate symplectic basis takes the following form:

$$\mathcal{R}(D\tau) = De, \qquad \mathcal{R}(JD\tau N^{-1}) = D\tau_+\,S + E, \tag{26}$$

where the matrix S is the one defined in (14) and the size of the matrix E is controlled in terms of the size of De (see formula (88) and related computations). Let us briefly explain the geometric and dynamical meaning of equations in (26). Suppose for a moment that $\mathcal{T}$ is an invariant torus and that τ is a solution of (8). In this case, we have that $e = 0$, $\Omega = 0$, and $E = 0$. Consequently, the columns of the matrices $JD\tau N^{-1}$ and $D\tau$ form a symplectic basis at every point of $\mathcal{T}$ and equations in (26) can be re-writen as follows:

$$\mathcal{R}(D\tau) = 0, \qquad \mathcal{R}(JD\tau N^{-1} + D\tau\,A) = D\tau_+\,\langle S\rangle_\theta, \tag{27}$$

where the matrix A is selected so that $\mathcal{L}_\omega A = S - \langle S\rangle_\theta$. Let us consider the linear quasi-periodic cocycle $(\theta, z) \in \mathbb{T}^r \times \mathbb{R}^{2r} \mapsto (\theta + \omega, DF(\tau(\theta))\,z)$, which is obtained by linearization of F around τ. Expressions in (27) mean that this linear map can be reduced to constant coefficients by a linear quasi-periodic transformation of the form $z = \Phi(\theta)Z$ (in the aim of the quasi-periodic Floquet theory for maps). The columns of the matrix Φ can be selected as the columns of the matrices $JD\tau N^{-1} + D\tau\,A$ and $D\tau$, and the reduced cocycle is of the form $(\theta, Z) \mapsto (\theta + \omega, \mathcal{S}\,Z)$, where the constant matrix $\mathcal{S}$ is defined as $\mathcal{S} = \begin{pmatrix} \mathrm{Id}_r & 0 \\ \langle S\rangle_\theta & \mathrm{Id}_r \end{pmatrix}$. This fact is referred to as the automatic reducibility of quasi-periodic Lagrangian tori of exact symplectic maps. Expressions in (26) guarantee the automatic quasi-reducibility of the quasi-torus $\mathcal{T}$ if $\|e\|_\rho \ll 1$ (i.e., reducibility except by a small error term). Both aspects above, approximate Lagrangian character and automatic quasi-reducibility, constitute the cornerstone of the method introduced in [7], as well as of our modified approach to the parameterization method. Indeed, if $\mathcal{T}$ is a quasi-torus, then we can rewrite equation (22) as follows (modulo quadratic terms on e, De, E, a, and b that we remove from the equation):

$$JD\tau_+ N_+^{-1}\,\mathcal{L}_\omega\,a + D\tau_+\,[\mathcal{L}_\omega\,b - S\,a] = e.$$

We multiply this equation by $(D\tau_+)^\top J$ and $(D\tau_+)^\top$, and remove the quadratic error terms. Then, we derive the (cohomological) equations for a and b considered in [7]:

$$\mathcal{L}_\omega a = -(D\tau_+)^\top J\,e, \qquad \mathcal{L}_\omega b = S\,a + N_+^{-1}(D\tau_+)^\top\,e. \tag{28}$$

Since ω is Diophantine, this triangular system of equations can be solved provided that $\langle(D\tau_+)^\top J\,e\rangle_\theta = 0$ and that $\det(\langle S\rangle_\theta) \neq 0$. Under these assumptions, the solutions for a and b of (28) are unique except by the value of

$\langle b \rangle_\theta$, that it is free to take any value (we may set $\langle b \rangle_\theta = 0$). Although the condition $\langle (D\tau_+)^\top J e \rangle_\theta = 0$ is not satisfied in general, in [7] it is shown that this average has quadratic size in terms of e and De. Explicitly, we have that $\langle (D\tau_+)^\top J e \rangle_\theta = -\frac{1}{2} \langle (De)^\top J e \rangle_\theta$ (e.g., compute the average of formula (83) and use that $\langle e_2 \rangle_\theta = 0$, as shown in remark 3.4). Therefore, if we replace the first equation in (28) by $\mathcal{L}_\omega a = -(D\tau_+)^\top J e + \langle (D\tau_+)^\top J e \rangle_\theta$, and we solve the new system of equations thus obtained, we still have a quasi-Newton method for equation (8).

In [7], it is proved the convergence of this quasi-Newton method if $\gamma^{-4} \rho^{-4\nu} \|e\|_\rho$ is sufficiently small (γ and ν are those in (3)). Our main purpose is to develop a new approach to the parameterization method for which the convergence condition is satisfied when the initial error is of $\mathcal{O}(\gamma^2)$. We also pay attention to the exponent of ρ in the convergence condition, since it plays a crucial role if one plans to extend the result to finitely differentiable maps (see remark 2.9). To accomplish these aims, we have adapted to exact symplectic maps the construction we introduced in [31] to compute Lagrangian tori of Hamiltonian systems. The main idea that we collect from [31] is that establishing convergence if $\|e\|_\rho$ is of $\mathcal{O}(\gamma^2)$ appears to be a very difficult task if we deal with the invariance error e as a single source of error for the parameterization τ. Indeed, it seems necessary to separately control the size of some projections of the invariance error, being the most important aspect the control of the distance of $\mathcal{T}$ to a Lagrangian manifold. As noted above, in [7] this distance is controled by bounding the size of Ω in terms of the size of De by formula (23) (see proposition 2.5). In this way, we can establish 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. To construct our modified parameterization method, we rely on the characterization of invariant Lagrangian tori of exact symplectic maps provided by proposition 3.2 below. The three equations for τ introduced in (29) are inspired by the three equations established in proposition 3.2 of [31] for Lagrangian tori of Hamiltonian systems. Although we have in mind the context of analytic functions, this result holds under mild regularity conditions.

Proposition 3.2. *With the same notations introduced in definitions 2.1, 2.2, and 2.3. Suppose F is an exact symplectic map of $\mathbb{R}^{2r}$ and $\omega \in \mathbb{R}^r$ is a (rotation) vector that verifies the non-resonance conditions $\langle k, \omega \rangle \neq m$, $\forall k \in \mathbb{Z}^r \setminus \{0\}$ and $\forall m \in \mathbb{Z}$. Let τ be the parameterization of a torus $\mathcal{T} \subset \mathbb{R}^{2r}$, of dimension r, for which the following conditions hold:*

$$(\mathrm{eq}_1)\, \tau_+^\top J\, F(\tau) = 0, \qquad (\mathrm{eq}_2)\, \mathcal{L}_\omega \vartheta - \nabla(V(\tau)) = 0, \qquad (\mathrm{eq}_3)\, (D\tau_+)^\top (F(\tau) - \tau_+) = 0. \qquad (29)$$

We also suppose that the Jacobian matrix of the invariance error e associated to F, τ, and ω (see (21)) is small enough so that $\|(De)^\top D\tau_+ N_+^{-1}\|_{\mathbb{T}^r} < 2$. Then, $\mathcal{T}$ is a Lagrangian torus and τ is a solution of the invariance equation (8).

Remark 3.3. *If we use proposition 3.2 to characterize invariant tori, the condition $\|(De)^\top D\tau_+ N_+^{-1}\|_{\mathbb{T}^r} < 2$ does not mean any strong constraint for the parameterization τ. Furthermore, the rotation vector ω should be nonresonant, but not necessarily Diophantine. We note that (eq_1) can also be written as $\tau_+^\top J e = 0$. Moreover, the result of proposition 3.2 is also true if $\tau_+^\top J\, F(\tau)$ is a constant function, but not necessarily zero. However, since proposition 3.2 implies $e = 0$, we can conclude a posteriori that this constant must be zero. Finally, using the definition of ϑ in (6) and formula (7), we can rewrite (eq_2) as:*

$$(\mathrm{eq}_2)\, (D\tau_+)^\top J\, \tau_+ = (D\tau)^\top (DF(\tau))^\top J\, F(\tau). \qquad (30)$$

Equation (30) has the (practical) advantage that does not involve the primitive function V.

Proof of proposition 3.2: We recall (see defintion 2.2) that ϑ is the vector representation of the one-form $\tau^* \varpi$, that Ω is the matrix representation of $\tau^* (dx \wedge dy)$, and that $d\varpi = -2\, dx \wedge dy$. Therefore, equation (eq_2) is equivalent to say that $(\tau_+)^* \varpi - \tau^* \varpi - d(V \circ \tau) = 0$. By taking the exterior derivative of this expression, we obtain $\mathcal{L}_\omega \Omega = 0$. Since the components of ω are nonresonant and $\langle \Omega \rangle_\theta = 0$, we conclude that $\Omega = 0$ (see equation (10)). Consequently, $\mathcal{T}$ is a Lagrangian torus and the columns of the matrices $JD\tau N^{-1}$ and $D\tau$ form

a symplectic basis, at any point of $\mathcal{T}$. We express the invariance error as $e = D\tau_+\, \alpha + JD\tau_+ N_+^{-1}\, \beta$, for some vector functions $\alpha, \beta : \mathbb{T}^r \to \mathbb{R}^r$. Equation (eq$_3$) implies that $(D\tau_+)^\top e = N_+\, \alpha = 0$. Hence, we have that $\alpha = 0$ and that $e = JD\tau_+ N_+^{-1}\, \beta$. Next, we use that equation (eq$_1$) means that $\tau_+^\top J\, e = 0$. By computing the gradient of this scalar expression, we obtain:

$$-(De)^\top J\, \tau_+ + (D\tau_+)^\top J\, e = 0. \tag{31}$$

By applying equations (21) and (24) to the right-hand side of the expression in (30), we obtain:

$$(D\tau_+)^\top J\, \tau_+ = (D\tau_+ + De)^\top J\, (\tau_+ + e).$$

As a result, we obtain the following relation:

$$(De)^\top J\, \tau_+ + (D\tau_+)^\top J\, e = -(De)^\top J\, e. \tag{32}$$

If we perform the sum of equations (31) and (32), and we use that $e = JD\tau_+ N_+^{-1}\, \beta$, then we derive the following equation for β:

$$\beta = -\frac{1}{2}\, (De)^\top D\tau_+\, N_+^{-1}\, \beta. \tag{33}$$

Equivalently, we have that β verifies $(\mathrm{Id}_r + \frac{1}{2}\, (De)^\top D\tau_+\, N_+^{-1})\, \beta = 0$. Since $\|(De)^\top D\tau_+ N_+^{-1}\|_{\mathbb{T}^r} < 2$, we have that $\det(\mathrm{Id}_r + \frac{1}{2}\, (De)^\top D\tau_+\, N_+^{-1})$ along the torus. Consequently, equation (33) implies that $\beta = 0$, as wanted.

$$\square$$

The conditions for τ defined by equations in (29) have a clear geometric and dynamical meaning, and are apparent if the torus is invariant. The scalar equation (eq$_1$) means that if we evaluate the one-form ϖ on the torus $\mathcal{T}$, using the parameterization τ_+, and we apply it to the invariance error e of (21), then the result is zero. Equation (eq$_2$) is due to the fact that $F^* \varpi = \varpi + dV$. If we perform the pull-back by τ of this formula, then we obtain that $(F \circ \tau)^* \varpi = \tau^* \varpi + d(V \circ \tau)$. If τ satisfies the invariance condition $F(\tau) = \tau_+$, then the coordinate representation of this identity is given by equation (eq$_2$). Finally, we justify equation (eq$_3$) as follows (we do not provide all the details because our approach does not use the construction below). Suppose that $\mathcal{T}$ is a Lagrangian torus (this fact is guaranteed by equation (eq$_2$) alone). If the invariance error e of τ is sufficiently small, then we have:

$$F(\tau(\theta)) = \tau(\theta + \omega + a(\theta)) + J\, D\tau(\theta + \omega + a(\theta))\, N^{-1}(\theta + \omega + a(\theta))\, b(\theta), \qquad \forall \theta \in \mathbb{T}^r, \tag{34}$$

for some small vector functions $a, b : \mathbb{T}^r \to \mathbb{R}^r$. These functions are uniquely defined by these conditions. This can be proved by the fixed point method applied to a functional $(a(\theta), b(\theta)) \mapsto \mathcal{F}(a(\theta), b(\theta))$. Indeed, if $\mathcal{G}(a(\theta), b(\theta))$ denotes the expression at the right-hand side of equation (34), then we can rewrite this equation as:

$$D\tau_+\, a + JD\tau_+\, N_+^{-1}\, b = e - (\mathcal{G}(a, b) - \tau_+ - D\tau_+\, a - JD\tau_+\, N_+^{-1}\, b). \tag{35}$$

Then, we define $\mathcal{F} = (\mathcal{F}_1, \mathcal{F}_2)$ by solving for $\mathcal{F}$ the linear system of equations obtained by replacing a by $\mathcal{F}_1$ and b by $\mathcal{F}_2$ at the left-hand side of equation (35). This expression for $\mathcal{F}$ is well-defined since the columns of $JD\tau N^{-1}$ and $D\tau$ form a basis at any point of $\mathcal{T}$. Once formula (34) has been set, we multiply it by $(D\tau(\theta + \omega + a(\theta)))^\top$. Then, the symplectic properties of $\mathcal{T}$ imply the following equation for a:

$$(D\tau(\theta + \omega + a(\theta)))^\top (F(\tau(\theta)) - \tau(\theta + \omega + a(\theta))) = 0. \tag{36}$$

Once again, if e is sufficiently small, then equation (36) also has a unique small solution for a. As done above for equation (34), this result may be proved by the fixed point method applied to a functional defined by the linearization of equation (36) (using now that $\det(N(\theta)) \neq 0$). Once established the uniqueness of solutions of equations (34) and (36), this means that the function a should be the same in both expressions. As a summary, if

$\mathcal{T}$ is a Lagrangian torus for which the invariance error e is sufficiently small, then equation (36) can be interpreted as a way to assign to $\mathcal{T}$ a sort of "inner dynamics" by the action of the map F, regardless of whether $\mathcal{T}$ is invariant or not. Therefore, equation (eq_2) means that if we compute a by equation (36), then we are imposing $a = \omega$. This is to say that this "pseudo-inner dynamics" for $\mathcal{T}$, using the parameterization τ, is the quasi-periodic rotation T_ω.

Proposition 3.2 means that we can address the computation of an invariant torus of F by solving for τ the nonlinear system of equations arising from the three conditions in (29) (see also equation (30)). Our initial aim is to solve this system by a quasi-Newton method. Therefore, given the parameterization τ of a quasi-torus of F, we must deal separately with three different error functions for τ. Explicitly:

$$e_1 = \tau_+^\top J F(\tau), \quad e_2 = (D\tau_+)^\top J \tau_+ - (D\tau)^\top (DF(\tau))^\top J F(\tau), \quad e_3 = (D\tau_+)^\top (F(\tau) - \tau_+). \tag{37}$$

Remark 3.4. *By remark 3.3, to apply proposition 3.2 we do not need to control the actual value of $\langle e_1 \rangle_\theta$. Indeed, what we get from this result is that if $\nabla e_1 = e_2 = e_3 = 0$, and $\|De\|_{\mathbb{T}^r}$ is sufficiently small, then $e = 0$ (and therefore $e_1 = 0$). Since e_2 can be written as $e_2 = \mathcal{L}_\omega \vartheta - \nabla(V(\tau))$, we have that $\langle e_2 \rangle_\theta = 0$. Moreover, using this formula for e_2 and definition 2.2, we conclude the following relation between Ω and e_2:*

$$\mathcal{L}_\omega \Omega = \frac{1}{2}(De_2 - (De_2)^\top). \tag{38}$$

Equation (38) *plays a crucial role in the proof of theorem 2.6. We also note that we can easily relate $\{e_j\}_{j=1}^3$ with e. By definition, $e_1 = \tau_+^\top J e$ and $e_3 = (D\tau_+)^\top e$. Formulas in* (80) *and* (81) *provide the expressions of ∇e_1 and e_2 in terms of e and De. We can also control the size of e in terms of the size of $\{e_j\}_{j=1}^3$, using the additional assumption that De is sufficiently small. This is done in the proof of theorem 2.6 (see formulas from* (79) *to* (87)*).*

Remark 3.5. *In [31], we used the fact that any Lagrangian torus that belongs to an energy level set of a Hamiltonian h is invariant by the flow of h (regardless of the inner dynamics of the torus). The scalar condition $e_1 = 0$ of proposition 3.2 (or the equivalent one $e_1 = \text{const.}$) plays a similar role in the present context, although we cannot assert that any Lagrangian torus that verifies $e_1 = 0$ is automatically invariant by the map F. Suppose for a moment that F is an exact symplectic diffeomorphism of the annulus $(\theta, I) \in \mathbb{T}^r \times \mathbb{R}^r$ and that $\mathcal{T}$ is a Lagrangian torus which is a graph over the angles, i.e., $\mathcal{T}$ can be parameterized as $I = \theta + \nabla\alpha(\theta)$, where $\alpha(\theta)$ is a scalar function. Then, we can introduce a specific "energy function", defined in terms of α and the primitive function of F, so that it takes a constant value on $\mathcal{T}$ if it is invariant by F. This idea is used in [12] to introduce a nonlinear system of two equations that characterize the Lagrangian tori that are invariant by F, carrying quasi-periodic motion with a prefixed rotation vector, and which are also a graph over θ. The way in which this nonlinear system can be solved by a quasi-Newton method is also schematized in [12] (although convergence details are not discussed).*

Suppose τ is the the parameterization of a quasi-torus of F for which ∇e_1, e_2, and e_3 are small vector functions. We also assume that the non-degeneracy twist condition $\det(\langle S \rangle_\theta) \neq 0$ holds for τ, where the matrix S is defined in (14). We seek a new parameterization, $\tau^{(1)}$, for which the corresponding errors behave almost quadratically with respect to the previous ones. The new errors, $\{e_j^{(1)}\}_{j=1}^3$, are defined by replacing τ by $\tau^{(1)}$ in (37). In definition 3.6 we formulate the quasi-Newton-like method we use to solve equations in (29) (compare with remark 3.4 of [31]).

Definition 3.6 (The modified parameterization method). *With notations above, we introduce $\tau^{(1)}$ through the following construction:*

$$\tau^{(1)}(\theta) = \bar{\tau}^{(1)}(\theta + b(\theta)), \tag{39}$$

where

$$\bar{\tau}^{(1)} = \tau + \Delta\tau, \qquad \Delta\tau = JD\tau \, N^{-1} a, \qquad \langle b \rangle_\theta = 0, \tag{40}$$

and the (small) vector functions $a(\theta)$ and $b(\theta)$ are defined by the following "generalized cohomolgical equations":

$$\mathcal{L}_\omega a = \frac{1}{2}e_2 - \frac{1}{2}\nabla e_1, \qquad \mathcal{L}_\omega b = S(\theta + b)\, a(\theta + b) + N^{-1}(\theta + \omega + b)\, e_3(\theta + b). \tag{41}$$

Explicitly, we solve equations in (41) *as follows. Firstly, we write* a *as* $a = \langle a \rangle_\theta + \widetilde{a}$, *where* $\widetilde{a}$ *is computed as:*

$$\widetilde{a} = \mathcal{L}_\omega^{-1} \left(\frac{1}{2} e_2 - \frac{1}{2} \nabla e_1 \right). \tag{42}$$

Next, we compute $\xi = \langle a \rangle_\theta$ *and* b *as the limit of the sequences* $\{\xi^{(n)}\}_{n \geq 0}$ *and* $\{b^{(n)}\}_{n \geq 0}$ *introduced by the iterative scheme below, starting with* $\xi^{(0)} = 0$ *and* $b^{(0)} = 0$. *Suppose* $\xi^{(n)}$ *and* $b^{(n)}$ *are known, for some* $n \geq 0$. *Then, we compute* $\xi^{(n+1)} = \xi^{(n)} + \Delta \xi^{(n)}$ *and* $b^{(n+1)} = b^{(n)} + \Delta b^{(n)}$, *with* $\langle \Delta b^{(n)} \rangle_\theta = 0$, *by the following construction:*

$$\eta^{(n)} = S \left[\widetilde{a} + \xi^{(n)} \right] + N_+^{-1} e_3, \quad E^{(n)} = \mathcal{L}_\omega b^{(n)} - \eta^{(n)} (\theta + b^{(n)}), \quad \mathcal{S}^{(n)} = (\mathrm{Id}_r + Db_+^{(n)})^{-1} S(\theta + b^{(n)}),$$

$$\Delta \xi^{(n)} = (\langle \mathcal{S}^{(n)} \rangle_\theta)^{-1} \langle (\mathrm{Id}_r + Db_+^{(n)})^{-1} E^{(n)} \rangle_\theta, \quad \zeta^{(n)} = \mathcal{L}_\omega^{-1} \left((\mathrm{Id}_r + Db_+^{(n)})^{-1} \left[S(\theta + b^{(n)}) \Delta \xi^{(n)} - E^{(n)} \right] \right),$$

$$\Delta \hat{b}^{(n)} = -\langle Db^{(n)} \zeta^{(n)} \rangle_\theta + \zeta^{(n)}, \quad \Delta b^{(n)} = (\mathrm{Id}_r + Db^{(n)}) \Delta \hat{b}^{(n)},$$

where we are assuming that $\det(\langle \mathcal{S}^{(n)} \rangle_\theta) \neq 0$ *and that* $\mathrm{Id}_r + Db^{(n)}$ *is non-singular along the torus (since* $\mathcal{S}^{(n)}$ *and* $\mathrm{Id}_r + Db^{(n)}$ *are "small perturbations" of* S *and* Id_r).

Therefore, to perform definition 3.6, we have to solve the two small-divisors equations in (41). The first one is a linear equation for a (in fact, for $\widetilde{a} = a - \langle a \rangle_\theta$), but the second one is a non-linear equation for $\xi = \langle a \rangle_\theta$ and b. As pointed out in the introduction, the nonlinear character of this second equation is the reason why we are referring to them as "generalized cohomological equations" (compare the second equation in (41) with its linearized version in (45)). Since $\langle e_2 \rangle_\theta = 0$ (see remark 3.4), the functions are analytic and the rotation vector ω is Diophantine, then the first equation in (41) has a unique solution for $\widetilde{a}$ (given by formula (42)), while ξ is free to take any value. The solutions for ξ and b of the second equation in (41) are constructed through an (explicit) iterative scheme, in which the hypothesis $\det(\langle S \rangle_\theta) \neq 0$ plays a crucial role. Below, in proposition 3.7, we show that this iterative scheme constitutes a quasi-Newton method for this nonlinear equation and that it is convergent if ∇e_1, e_2, and e_3 are sufficiently small. Before doing this, we show in a schematic way that $\tau^{(1)}$ in (39) is defined through a quasi-Newton-like method for equations in (29) and we deduce equations in (41). To do this, we start by looking for the new parameterization $\tau^{(1)}$ as $\tau^{(1)} = \tau + JD\tau N^{-1} a + D\tau b$. We compute the linear approximations around τ of the new errors $\{e_j^{(1)}\}_{j=1}^3$ associated to this simplified expression for $\tau^{(1)}$, and we equate to zero the resulting expressions. After some computations, we obtain the following (linear) equations for a and b:

$$\tau_+^\top D\tau_+ N_+^{-1} \mathcal{L}_\omega a = -e_1 + \tau_+^\top J D\tau_+ N_+^{-1} e_3, \tag{43}$$

$$\mathcal{L}_\omega a = \frac{1}{2} e_2 - \frac{1}{2} \nabla e_1, \tag{44}$$

$$\mathcal{L}_\omega b - S a = N_+^{-1} e_3. \tag{45}$$

Computations leading to equations above are analogous to those performed in the proof of theorem 2.6, so we postpone the details until section 4. In particular, these computations involve the automatic quasi-reducibility of a quasi-torus (see equation (88) and related formulas). It is easy to realize that, under the same assumptions of definition 3.6, equations (44) and (45) provide unique solutions for a and b (if we set $\langle b \rangle_\theta = 0$). Once a and b have been defined in this way, some straight computations show that equation (43) is fulfilled automatically, except by a quadratic error term (again we refer to section 4 for the details). Therefore, we can build a quasi-Newton method for equations in (29) by computing $\tau^{(1)} = \tau + JD\tau N^{-1} a + D\tau b$ (as done in [7], see equation (25)), with a and b defined by equations (44) and (45) (the cohomological equations for a and b are different from those in [7], see (28)). However, proceeding in this way we have not been able to improve the error estimates of the original approach to the parameterization method of [7]. Roughly speaking, the reason of the failure of this modified

approach is that the errors $\{e_j^{(1)}\}_{j=1}^3$ associated to this simplified expression of $\tau^{(1)}$ contain some terms whose size do not behave appropriately. More specifically these terms are due to the contribution of $D\tau\,b$ to $\tau^{(1)}$.

To try to overcome this drawback, we rely on the same solution adopted to carry out the modified parameterization method for Hamiltonian systems in [31]. This solution consists in using formulas as compact as possible for the involved objects. Explicitly, this compactness is achieved by introducing the contribution of b to $\tau^{(1)}$ through a composition, rather than by an addition (see equations (39) and (40)). Proceeding in this way, the results obtained in [31] were satisfactory, i.e., the size of the new errors behave appropriately. However, within the present context of maps, considering a compact expression for $\tau^{(1)}$ seems not to be enough to solve the problem. Indeed, if $\tau^{(1)}$ is defined by equations (39) and (40), but a and b are the solutions of the linear (cohomological) equations (44) and (45), then the new errors still behave in a similar way as those in [7]. The reason for this behavior is that equation (45) does not reflect the fact that b enters in $\tau^{(1)}$ through a composition. We overcome this additional drawback by replacing the linear equation (45) by the second (non-linear) equation in (41).

Although the selected construction involves a non-linear equation for b, it can be solved iteratively in a fairly simple way by the quasi-Newton method introduced at the bottom part of definition 3.6 (see the introduction for additional comments regarding to the resolution of this equation). The convergence of this quasi-Newton method is discussed below. Indeed, in proposition 3.7 we construct solutions for $\langle a \rangle_\theta$ and b (with $\langle b \rangle_\theta = 0$) of the second equation in (41), with the understanding that we have already computed $\widetilde{a} = a - \langle a \rangle_\theta$ by formula (42). The quantitative estimates for $\langle a \rangle_\theta$ and b provided by this result are similar to those that would result for these objects as solutions of the linear equation (45) (we leave the details for the interested reader). Furthermore, the conditions on the smallness of $\{e_j\}_{j=1}^3$ required to apply proposition 3.7 are compatible with those we establish to ensure the convergence of the iterative application of the parameteritzation method introduced in definition 3.6 (see the proof of theorem 2.6 for details). Although the uniqueness of solution of this non-linear equation is not discussed in proposition 3.7, it is natural to expect that we have unique solutions for $\langle a \rangle_\theta$ and b if only "small solutions" of this equation are taken into account. But we do not mind about the uniqueness of this solution, since it is not necessary for our construction. Actually, if we are interested on dealing with a problem depending on parameters in a differentiable way, then this regulatity is preserved by the solutions for $\langle a \rangle_\theta$ and b provided by proposition 3.7. In the statement of proposition 3.7 we basically keep the same names introduced above for the objects involved. We only introduce the compact notations $\xi = \langle a \rangle_\theta$ and $\hat{e} = N_+^{-1} e_3$. Furthermore, hypotheses of the statement are coherent with the bounds that we have for the involved objects along the proof of theorem 2.6 (see section 4).

Proposition 3.7. *With the notations and definitions introduced before. Let $\widetilde{a}, \hat{e} : \mathbb{T}^r \to \mathbb{R}^r$ and $S : \mathbb{T}^r \to \mathbb{M}_{r \times r}(\mathbb{R})$ be real analytic functions, verifying $\langle \widetilde{a} \rangle_\theta = 0$ and $\det(\langle S \rangle_\theta) \neq 0$. We suppose that all these functions can be analytically extended to $\Delta(\bar{\rho})$, for some $\bar{\rho} > 0$, and that the following estimates hold for some constants $m \geq 1$, $0 < \delta < \min\{1, \bar{\rho}/4\}$, and $0 < \mu < 1$:*

$$\|f\|_{\bar{\rho}} \leq m\mu, \qquad \|Df\|_{\bar{\rho}} \leq m\frac{\mu}{\delta}, \qquad \|D^2 f\|_{\bar{\rho}} \leq m\frac{\mu}{\delta^2}, \tag{46}$$

where the function f above stands for any component of the r-dimensional vector functions $S\,\widetilde{a}$ and $\hat{e}$, and:

$$|(\langle S \rangle_\theta)^{-1}| \leq m, \quad \|S\|_{\bar{\rho}} \leq m, \quad \|DS_{j,l}\|_{\bar{\rho}} \leq \frac{m}{\delta}, \quad \|D^2 S_{j,l}\|_{\bar{\rho}} \leq \frac{m}{\delta^2}, \quad \forall j, l = 1, \ldots, r, \tag{47}$$

where $S = (S_{j,l})_{j,l=1,\ldots,r}$ are the components of the matrix S. We also consider a (rotation) vector $\omega \in \mathbb{R}^r$ that satisfies the Diophantine conditions (3), for some $0 < \gamma \leq 1$ and $\nu \geq r$, and we denote by $\sigma = \sigma(r, \nu) \geq 1$ the constant provided by proposition 2.5. There is a constant $\bar{m} \geq m$, that depends only on r, σ, and m, for which the following result holds. Suppose that μ is sufficiently small so that:

$$2^{2+2\nu} r \sigma m \bar{m}^2 \frac{\mu}{\gamma \delta^{\nu+1}} \leq 1. \tag{48}$$

Then, there are $\xi \in \mathbb{R}^r$ and a real analytic function $b : \mathbb{T}^r \to \mathbb{R}^r$, with $\langle b \rangle_\theta = 0$, in such a way $a = \xi + \widetilde{a}$ and b satisfy the following equation:

$$\mathcal{L}_\omega b = S(\theta + b) \, a(\theta + b) + \hat{e}(\theta + b). \tag{49}$$

Moreover, the function b can be analytically extended to $\Delta(\bar{\rho} - 4\delta)$ and the following estimates hold:

$$|\xi| \leq 4m^2\mu, \qquad \|b\|_{\bar{\rho}-4\delta} \leq 8\sigma m^3 \frac{\mu}{\gamma\delta^\nu}, \qquad \|Db\|_{\bar{\rho}-4\delta} \leq 8r\sigma m^3 \frac{\mu}{\gamma\delta^{\nu+1}}. \tag{50}$$

To perform the proof of proposition 3.7 (and later on of theorem 2.6), we use some basic properties of the norm $\|\cdot\|_\rho$ (see definition 2.4). Explicitly, if f is a function defined in $\mathbb{T}^r$ (scalar-valued, vector-valued or matrix-valued), that can be analytically extended to the complex strip $\Delta(\rho)$ and is bounded in the closure of $\Delta(\rho)$, then we have the following bounds for the average of f, for $\widetilde{f} = f - \langle f \rangle_\theta$, and for the derivatives of f (Cauchy estimates):

$$|\langle f \rangle_\theta| \leq \|f\|_\rho, \qquad \|\widetilde{f}\|_\rho \leq 2\|f\|_\rho, \qquad \|\partial_{\theta_j} f\|_{\rho-\delta} \leq \frac{\|f\|_\rho}{\delta}, \quad j = 1, \ldots, r. \tag{51}$$

Let f_1 and f_2 be two matrix-valued functions with the same analytic properties as f above. If the product $f_1 f_2$ is well-defined, then $\|\cdot\|_\rho$ verifies the multiplicative property $\|f_1 f_2\|_\rho \leq \|f_1\|_\rho \cdot \|f_2\|_\rho$. Moreover, if $A \in \mathbb{M}_{r \times r}(\mathbb{C})$ satisfies $|A| < 1$, then we use Neumann's series to control the inverse of a matrix $\mathrm{Id}_r + A$. We obtain:

$$|(\mathrm{Id}_r + A)^{-1}| \leq \frac{1}{1 - |A|}, \qquad |(\mathrm{Id}_r + A)^{-1} - \mathrm{Id}| \leq \frac{|A|}{1 - |A|}. \tag{52}$$

By the multiplicative property of the norm $\|\cdot\|_\rho$, expressions in (52) can be easily extended to it.

Proof of proposition 3.7: We compute the solutions for b and ξ iteratively, as the limit of sequences $\{\xi^{(n)}\}_{n \geq 0}$ and $\{b^{(n)}\}_{n \geq 0}$. We introduce these sequences by the quasi-Newton method that we have schematized at the bottom part of definition 3.6 (recall that $\hat{e} = N_+^{-1} e_3$). Firstly, we discuss the formal aspects of this iterative scheme, without paying attention to the conditions that guarantee the well-defined character of the related expressions (mainly of the compositions involved). Subsequently, we prove the convergence of the iterative application of this scheme.

Explicit expressions of one step of the iterative scheme. We start with $b^{(0)} = 0$ and $\xi^{(0)} = 0$. Suppose we have computed the iterates $b^{(n)}$ and $\xi^{(n)}$, for some $n \geq 0$, and denote by $E^{(n)}$ the corresponding error function:

$$E^{(n)} = \mathcal{L}_\omega b^{(n)} - \eta^{(n)}(\theta + b^{(n)}), \tag{53}$$

where

$$\eta^{(n)} = S[\widetilde{a} + \xi^{(n)}] + \hat{e}. \tag{54}$$

Solving equation (49) can be formulated in terms of the convergence of the sequences $\{\xi^{(n)}\}_{n \geq 0}$ and $\{b^{(n)}\}_{n \geq 0}$, by showing simultaneously that $E^{(n)}$ goes to zero. By taking partial derivatives in equation (53), we obtain:

$$\mathcal{L}_\omega (Db^{(n)}) = D\eta^{(n)}(\theta + b^{(n)}) (\mathrm{Id}_r + Db^{(n)}) + DE^{(n)}. \tag{55}$$

Relation above plays a crucial role in the proof. We express $b^{(n+1)}$ and $\xi^{(n+1)}$ as follows:

$$b^{(n+1)} = b^{(n)} + \Delta b^{(n)}, \qquad \xi^{(n+1)} = \xi^{(n)} + \Delta \xi^{(n)},$$

where $\Delta b^{(n)}$ and $\Delta \xi^{(n)}$ are small corrections to be computed. After iteration, the new error function is:

$$\begin{aligned} E^{(n+1)} &= \mathcal{L}_\omega (b^{(n+1)}) - \eta^{(n+1)}(\theta + b^{(n+1)}) \\ &= E^{(n)} + \mathcal{L}_\omega (\Delta b^{(n)}) + \eta^{(n)}(\theta + b^{(n)}) - \eta^{(n)}(\theta + b^{(n+1)}) - S(\theta + b^{(n+1)}) \Delta\xi^{(n)} \\ &= E^{(n)} + \mathcal{L}_\omega (\Delta b^{(n)}) - D\eta^{(n)}(\theta + b^{(n)}) \Delta b^{(n)} - S(\theta + b^{(n)}) \Delta\xi^{(n)} - \bar{E}^{(n+1)} - \hat{E}^{(n+1)} \Delta\xi^{(n)}, \end{aligned} \tag{56, 57}$$

where

$$\begin{aligned}
\bar{E}^{(n+1)} &= \eta^{(n)}(\theta + b^{(n+1)}) - \eta^{(n)}(\theta + b^{(n)}) - D\eta^{(n)}(\theta + b^{(n)})\,\Delta b^{(n)}, \\
\hat{E}^{(n+1)} &= S(\theta + b^{(n+1)}) - S(\theta + b^{(n)}).
\end{aligned}$$

To control $\bar{E}^{(n+1)}$ and $\hat{E}^{(n+1)}$, we consider the following integral expressions for their components, which are straightforward consequence of the integral form of the remainder of Taylor's formula:

$$\bar{E}_j^{(n+1)} = (\Delta b^{(n)})^\top \left[\int_0^1 (1-s)\,D^2\eta_j^{(n)}(\theta + b^{(n)} + s\Delta b^{(n)})\,ds \right] \Delta b^{(n)}, \quad j = 1,\ldots,r, \tag{58}$$

$$\hat{E}_{j,l}^{(n+1)} = \left[\int_0^1 DS_{j,l}(\theta + b^{(n)} + s\Delta b^{(n)})\,ds \right] \Delta b^{(n)}, \quad j,l = 1,\ldots,r. \tag{59}$$

Suppose for a moment that $\mathrm{Id}_r + Db^{(n)}$ is non-singular along the torus. Then, we rewrite $\Delta b^{(n)}$ as follows:

$$\Delta b^{(n)} = (\mathrm{Id}_r + Db^{(n)})\,\Delta\hat{b}^{(n)}, \tag{60}$$

where $\Delta\hat{b}^{(n)}$ is the new unknown. In particular, we have:

$$\mathcal{L}_\omega\left(\Delta b^{(n)}\right) = \mathcal{L}_\omega\left(Db^{(n)}\right)\Delta\hat{b}^{(n)} + (\mathrm{Id}_r + Db_+^{(n)})\,\mathcal{L}_\omega\left(\Delta\hat{b}^{(n)}\right). \tag{61}$$

We replace $\mathcal{L}_\omega\left(\Delta b^{(n)}\right)$ by formula in (61) and $\mathcal{L}_\omega\left(Db^{(n)}\right)$ by formula in (55) in the expression (57) for $E^{(n+1)}$. Since we want the size of $E^{(n+1)}$ to be "quadratic", we consider the following equation for $\Delta\hat{b}^{(n)}$ and $\Delta\xi^{(n)}$, which cancels out the "non-quadratic" terms of the resulting expression:

$$\mathcal{L}_\omega\left(\Delta\hat{b}^{(n)}\right) = (\mathrm{Id}_r + Db_+^{(n)})^{-1}\left[S(\theta + b^{(n)})\,\Delta\xi^{(n)} - E^{(n)} \right]. \tag{62}$$

We solve equation (62) as follows. Firstly, we select $\Delta\xi^{(n)}$ to cancel out the average with respect to θ of the right-hand side of (62). Explicitly, we introduce the matrix

$$\mathcal{S}^{(n)} = (\mathrm{Id}_r + Db_+^{(n)})^{-1}\,S(\theta + b^{(n)}), \tag{63}$$

and we suppose that $b^{(n)}$ is sufficiently small so that the constant matrix $\langle\mathcal{S}^{(n)}\rangle_\theta$ is non-singular. Then, we obtain:

$$\Delta\xi^{(n)} = (\langle\mathcal{S}^{(n)}\rangle_\theta)^{-1}\,\langle(\mathrm{Id}_r + Db_+^{(n)})^{-1}\,E^{(n)}\rangle_\theta.$$

Next, we compute $\zeta^{(n)} = \Delta\hat{b}^{(n)} - \langle\Delta\hat{b}^{(n)}\rangle_\theta$ by applying $\mathcal{L}_\omega^{-1}$ to the right-hand side of (62) (see equation (10)),

$$\zeta^{(n)} = \mathcal{L}_\omega^{-1}\left((\mathrm{Id}_r + Db_+^{(n)})^{-1}\left[S(\theta + b^{(n)})\,\Delta\xi^{(n)} - E^{(n)} \right] \right),$$

and we select $\langle\Delta\hat{b}^{(n)}\rangle_\theta$ so that the average of $\Delta b^{(n)}$ (see bequation (60)) is equal to zero, i.e.,

$$\langle\Delta\hat{b}^{(n)}\rangle_\theta = -\langle Db^{(n)}\,\zeta^{(n)}\rangle_\theta.$$

Computations above imply the following expression for $E^{(n+1)}$ (see (55), (56), (61), and (62)):

$$E^{(n+1)} = DE^{(n)}\,\Delta\hat{b}^{(n)} - \bar{E}^{(n+1)} - \hat{E}^{(n+1)}\,\Delta\xi^{(n)}. \tag{64}$$

Bounding the first step of the iterative scheme. Since $\xi^{(0)} = 0$ and $b^{(0)} = 0$, we obtain the following expressions:

$$E^{(0)} = -S\,\tilde{a} - \hat{e}, \quad \xi^{(1)} = \Delta\xi^{(0)} = (\langle S\rangle_\theta)^{-1}\,\langle E^{(0)}\rangle_\theta, \quad b^{(1)} = \Delta b^{(0)} = \mathcal{L}_\omega^{-1}(S\,\xi^{(1)} - E^{(0)}),$$

and we derive the following estimates (we use proposition 2.5 to control $\mathcal{L}_\omega^{-1}$):

$$\|E^{(0)}\|_{\bar{\rho}} \leq 2m\mu, \quad |\xi^{(1)}| \leq 2m^2\mu, \quad \|b^{(1)}\|_{\bar{\rho}-\delta} \leq 4\sigma m^3 \frac{\mu}{\gamma\delta^\nu}, \quad \|Db^{(1)}\|_{\bar{\rho}-2\delta} \leq 4r\sigma m^3 \frac{\mu}{\gamma\delta^{\nu+1}}. \qquad (65)$$

We use the bounds in (65) as motivation to establish the estimates in (67), which we expect to hold iteratively.

Iterative bounds. We introduce the following quantities, defined for any $n \geq 0$, where the value of $\bar{m}$ (the constant of the statement of the proposition) still has to be set:

$$\bar{\chi} = 2^{1+2\nu} m\bar{m}\frac{\mu}{\gamma\delta^{\nu+1}}, \quad \bar{\mu}^{(n)} = 2^{1-2\nu n} m\bar{\chi}^{2^n-1}\mu, \quad \bar{\rho}^{(0)} = \bar{\rho}, \quad \bar{\delta}^{(n)} = \frac{\delta}{2^n}, \quad \bar{\rho}^{(n+1)} = \bar{\rho}^{(n)} - 2\bar{\delta}^{(n)}. \qquad (66)$$

We observe that $\bar{\chi} \leq 1/2$ (see hypothesis (48)), that $\lim_{n\to\infty} \bar{\mu}^{(n)} = 0$, and that $\lim_{n\to\infty} \bar{\rho}^{(n)} = \bar{\rho} - 4\delta$. Suppose that the following bounds hold after n iterations, for some $n \geq 0$:

$$\|E^{(n)}\|_{\bar{\rho}^{(n)}} \leq \bar{\mu}^{(n)}, \quad |\xi^{(n)}| \leq 4m^2\mu, \quad \|b^{(n)}\|_{\bar{\rho}^{(n)}} \leq 8\sigma m^3 \frac{\mu}{\gamma\delta^\nu}, \quad \|Db^{(n)}\|_{\bar{\rho}^{(n)}} \leq 8r\sigma m^3 \frac{\mu}{\gamma\delta^{\nu+1}}. \qquad (67)$$

We also suppose that:

$$\|b^{(n)}\|_{\bar{\rho}^{(n)}} + \bar{m}\frac{\bar{\mu}^{(n)}}{\gamma(\bar{\delta}^{(n)})^\nu} \leq \delta. \qquad (68)$$

Since $\xi^{(0)} = 0$, $b^{(0)} = 0$, $\|E^{(0)}\|_{\bar{\rho}} \leq 2m\mu$ (see (65)), and $\bar{\mu}^{(0)} = 2m\mu$ (see (66)), it is clear that for $n = 0$ bounds in (67) are met. In fact, according to (65), bounds for $\xi^{(n)}$, $b^{(n)}$, and $Db^{(n)}$ in (67) are even pessimistic for the case $n = 1$. This is because when writing bounds in (67) we basically had in mind the cases where $n \geq 2$. However, we prefer to deal with expressions in (67) iteratively, for $n \geq 0$, since then we are not forced to control $E^{(1)}$ separately of the general case $E^{(n)}$, for $n \geq 2$. Of course, given that the estimates in (68) for n = 1 are also true, we will take them into account when proving bounds in (70) and (71) by induction. Finally, the bound in (68) is satisfied for $n = 0$, since $b^{(0)} = 0$, $\bar{\mu}^{(0)} = 2m\mu$, $\delta^{(0)} = \delta$, and $2m\bar{m}\mu/(\gamma\delta^\nu) \leq \delta$ (see hypothesis (48)).

Now, we proceed by induction. We are going to show that if bounds in (67) and (68) are fulfilled for some $n \geq 0$, then the bounds are also true if we replace n by $n + 1$. Specifically, our aim is to construct a constant $\bar{m}$ for which the result holds. To achieve this outcome, during the next sequence of bounds the value of $\bar{m}$ is redefined recursively to meet a finite number of conditions. At the end of this process, the *last* value of $\bar{m}$ obtained is that of the statement of the proposition. A crucial point in carrying out the induction process is to guarantee that the compositions of functions involving $\theta + sb^{(n)}$ are well defined, for any $s \in [0, 1]$. Indeed, we observe that:

$$(\cdot + sb^{(n)}(\cdot))(\Delta(\bar{\rho}^{(n)})) \subset \Delta(\bar{\rho}), \qquad \forall s \in [0, 1]. \qquad (69)$$

The inclusion (69) for $n = 0$ is straightforward, since $b^{(0)} = 0$ and $\bar{\rho}^{(n)} = \bar{\rho}$. If $n \geq 1$ and the bound in (68) holds, then $\|b^{(n)}\|_{\bar{\rho}^{(n)}} \leq \delta$. Consequently:

$$(\cdot + sb^{(n)}(\cdot))(\Delta(\bar{\rho}^{(n)})) \subset \Delta(\bar{\rho}^{(n)} + \delta) \subset \Delta(\bar{\rho}^{(1)} + \delta) = \Delta(\bar{\rho} - \delta) \subset \Delta(\bar{\rho}), \qquad \forall s \in [0, 1].$$

Regarding to the invertibility of $\mathrm{Id}_r + Db^{(n)}$ and $\langle \mathcal{S}^{(n)} \rangle_\theta$ (see (63)), we are going to show:

$$\|(\mathrm{Id}_r + Db^{(n)})^{-1}\|_{\bar{\rho}^{(n)}} \leq 2, \qquad |(\langle \mathcal{S}^{(n)} \rangle_\theta)^{-1}| \leq 2m. \qquad (70)$$

The first bound in (70) is straightforward by using Neumann's series, since (48) and (67) imply $\|Db^{(n)}\|_{\bar{\rho}^{(n)}} \leq 1/2$. To prove the second one, we start by expressing the components of $S(\theta + b^{(n)}) - S$ as follows:

$$S_{j,l}(\theta + b^{(n)}(\theta)) - S_{j,l}(\theta) = \left[\int_0^1 DS_{j,l}(\theta + sb^{(n)}) \, ds\right] b^{(n)}(\theta), \qquad \forall j, l = 1, \ldots, r.$$

We use the bounds for $DS_{j,l}$ in (47) and equation (69) to control the composition $DS_{j,l}(\theta + sb^{(n)})$. We obtain:

$$\|S(\theta + b^{(n)}) - S\|_{\bar{\rho}^{(n)}} \leq \bar{m}\frac{\mu}{\gamma\delta^{\nu+1}},$$

for a value of $\bar{m}$ properly defined. Then, we write $\mathcal{S}^{(n)} - S$ as follows:

$$\mathcal{S}^{(n)} - S = (\mathrm{Id}_r + Db_+^{(n)})^{-1}\left[S(\theta + b^{(n)}) - S - Db_+^{(n)}S\right].$$

Using expression above, we establish the following bounds (by redefining $\bar{m}$):

$$\|\mathcal{S}^{(n)} - S\|_{\bar{\rho}^{(n)}} \leq \bar{m}\frac{\mu}{\gamma\delta^{\nu+1}}, \quad |\langle\mathcal{S}^{(n)}\rangle_\theta - \langle S\rangle_\theta| \leq \bar{m}\frac{\mu}{\gamma\delta^{\nu+1}}, \quad |(\langle S\rangle_\theta)^{-1}(\langle\mathcal{S}^{(n)}\rangle_\theta - \langle S\rangle_\theta)| \leq \bar{m}\frac{\mu}{\gamma\delta^{\nu+1}}. \quad (71)$$

Next, we write:

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

Since $\bar{m}\mu/(\gamma\delta^{\nu+1}) \leq 1/2$ (see hypothesis (48)), from the last of the bounds of equation (71) we conclude that $\langle\mathcal{S}^{(n)}\rangle_\theta$ is non-singular and satisfies the estimate in (70) (once again we rely on Neumann's series).

Then, we obtain the following bounds (we use (69) for $s = 1$ to control the composition $S(\theta + b^{(n)})$ and we also recall that $\|Db^{(n)}\|_{\bar{\rho}^{(n)}} \leq 1/2$):

$$|\Delta\xi^{(n)}| \leq \bar{m}\bar{\mu}^{(n)}, \qquad \|\zeta^{(n)}\|_{\bar{\rho}^{(n)}-\bar{\delta}^{(n)}} \leq \bar{m}\frac{\bar{\mu}^{(n)}}{\gamma(\bar{\delta}^{(n)})^\nu}, \qquad |\langle\Delta\hat{b}^{(n)}\rangle_\theta| \leq \bar{m}\frac{\bar{\mu}^{(n)}}{\gamma(\bar{\delta}^{(n)})^\nu},$$

$$\|\Delta\hat{b}^{(n)}\|_{\bar{\rho}^{(n)}-\bar{\delta}^{(n)}} \leq \bar{m}\frac{\bar{\mu}^{(n)}}{\gamma(\bar{\delta}^{(n)})^\nu}, \qquad \|\Delta b^{(n)}\|_{\bar{\rho}^{(n)}-\bar{\delta}^{(n)}} \leq \bar{m}\frac{\bar{\mu}^{(n)}}{\gamma(\bar{\delta}^{(n)})^\nu}. \qquad (72)$$

The assumed bounds for S, $S\,\widetilde{a}$, and $\hat{e}$ in the statement of the proposition, together with the inductive bound for the constant vector $\xi^{(n)}$, imply the following estimates for the Hessian of any component of $\eta^{(n)}$ (see (54)):

$$\|D^2\eta_j^{(n)}\|_{\bar{\rho}} \leq \bar{m}\frac{\mu}{\delta^2}, \qquad j = 1, \ldots, r.$$

Moreover, the bound in (72) for $\|\Delta b^{(n)}\|_{\bar{\rho}^{(n)}-\bar{\delta}^{(n)}}$, together with the inductive hypothesis (68), imply that

$$(\cdot + b^{(n)}(\cdot) + s\Delta b^{(n)}(\cdot))(\Delta(\bar{\rho}^{(n)} - \bar{\delta}^{(n)})) \subset \Delta(\bar{\rho}^{(n)} - \bar{\delta}^{(n)} + \delta) \subset \Delta(\bar{\rho}), \qquad \forall s \in [0,1]. \qquad (73)$$

For $n = 0$ the proof of the second of these inclusions of sets is slightly different from the general case, but it is straightforward. We use (73) to control the compositions of functions involved in (58) and (59). We obtain:

$$\|\bar{E}^{(n+1)}\|_{\bar{\rho}^{(n)}-\bar{\delta}^{(n)}} \leq \bar{m}\frac{\mu(\bar{\mu}^{(n)})^2}{\gamma^2\delta^2(\bar{\delta}^{(n)})^{2\nu}}, \qquad \|\hat{E}^{(n+1)}\|_{\bar{\rho}^{(n)}-\bar{\delta}^{(n)}} \leq \bar{m}\frac{\bar{\mu}^{(n)}}{\gamma\delta(\bar{\delta}^{(n)})^\nu}.$$

Consequently (see (64)):

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

Since hypothesis (48) implies that $\mu/(\gamma\delta^{\nu+1}) \leq 1$, we have that the bound $\|E^{(n+1)}\|_{\bar{\rho}^{(n+1)}} \leq \bar{\mu}^{(n+1)}$ holds, provided that we define $\bar{\mu}^{(n+1)}$ as follows:

$$\bar{\mu}^{(n+1)} = 2^{2\nu n}\bar{m}\frac{(\bar{\mu}^{(n)})^2}{\gamma\delta^{\nu+1}}. \qquad (74)$$

At this point, we set the value for the constant $\bar{m}$ as one for which all the previous bounds are satisfied. Since $\bar{\mu}^{(0)} = 2m\mu$ (see (65)), it is easy to verify that the quantities $\{\bar{\mu}^{(n)}\}_{n\geq 0}$, defined by the recurrence in (74), are those given in equation (66). Since $\lim_{n\to\infty}\bar{\mu}^{(n)} = 0$, we conclude that $\lim_{n\to\infty}\|E^{(n)}\|_{\bar{\rho}^{(n)}} = 0$.

To finish the proof, it only remains to prove that the other bounds in (67), as well as the bound in (68), are also met when replacing n by $n+1$. To achieve this outcome, we should consider the following bounds:

$$\sum_{j=1}^{\infty}\bar{\mu}^{(j)} \leq 8m^2\bar{m}\frac{\mu^2}{\gamma\delta^{\nu+1}}, \qquad \sum_{j=1}^{\infty}\frac{\bar{\mu}^{(j)}}{(\bar{\delta}^{(j)})^{\nu}} \leq 2^{3+\nu}m^2\bar{m}\frac{\mu^2}{\gamma\delta^{2\nu+1}}, \qquad \sum_{j=1}^{\infty}\frac{\bar{\mu}^{(j)}}{(\bar{\delta}^{(j)})^{\nu+1}} \leq 2^{4+\nu}m^2\bar{m}\frac{\mu^2}{\gamma\delta^{2\nu+2}}.$$

The above inequalities are immediate controlling each of these three series by a geometric one of ratio $1/2$. To do that, we use that $\bar{\chi} \leq 1/2$ and that $\bar{\chi}^{2^j-1} \leq \bar{\chi}^j$. Consequently, we obtain:

$$|\xi^{(n+1)}| \leq |\xi^{(1)}| + \sum_{j=1}^{n}|\Delta\xi^{(j)}| \leq 2m^2\mu + \bar{m}\sum_{j=1}^{\infty}\bar{\mu}^{(j)} \leq 2m^2\mu + 8m^2\bar{m}^2\frac{\mu^2}{\gamma\delta^{\nu+1}},$$

$$\|b^{(n+1)}\|_{\bar{\rho}^{(n+1)}} + \bar{m}\frac{\bar{\mu}^{(n+1)}}{\gamma(\bar{\delta}^{(n+1)})^{\nu}} \leq \|b^{(1)}\|_{\bar{\rho}-\delta} + \sum_{j=1}^{n}\|\Delta b^{(j)}\|_{\bar{\rho}^{(j)}-\bar{\delta}^{(j)}} + \bar{m}\frac{\bar{\mu}^{(n+1)}}{\gamma(\bar{\delta}^{(n+1)})^{\nu}}$$

$$\leq 4\sigma m^3\frac{\mu}{\gamma\delta^{\nu}} + \frac{\bar{m}}{\gamma}\sum_{j=1}^{\infty}\frac{\bar{\mu}^{(j)}}{(\bar{\delta}^{(j)})^{\nu}} \leq 4\sigma m^3\frac{\mu}{\gamma\delta^{\nu}} + 2^{3+\nu}m^2\bar{m}^2\frac{\mu^2}{\gamma^2\delta^{2\nu+1}},$$

and

$$\|Db^{(n+1)}\|_{\bar{\rho}^{(n+1)}} \leq \|Db^{(1)}\|_{\bar{\rho}-2\delta} + \sum_{j=1}^{n}\|D(\Delta b^{(j)})\|_{\bar{\rho}^{(j)}-2\bar{\delta}^{(j)}}$$

$$\leq \|Db^{(1)}\|_{\bar{\rho}-2\delta} + r\sum_{j=1}^{n}\frac{\|\Delta b^{(j)}\|_{\bar{\rho}^{(j)}-\bar{\delta}^{(j)}}}{\bar{\delta}^{(j)}}$$

$$\leq 4r\sigma m^3\frac{\mu}{\gamma\delta^{\nu+1}} + \frac{r\bar{m}}{\gamma}\sum_{j=1}^{\infty}\frac{\bar{\mu}^{(j)}}{(\bar{\delta}^{(j)})^{\nu+1}} \leq 4r\sigma m^3\frac{\mu}{\gamma\delta^{\nu+1}} + 2^{4+\nu}rm^2\bar{m}^2\frac{\mu^2}{\gamma^2\delta^{2\nu+2}}.$$

Hence, the desired bounds for the case $n+1$ are straightforward from hypothesis (48).

$\square$

4 Proof of the main result

We prove theorem 2.6 by the iterative application of the quasi-Newton-like method formulated in definition 3.6. We take $\tau^{(0)} = \tau$ as the initial guess and we generate a sequence $\{\tau^{(n)}\}_{n\geq 0}$ of parameterizations. Our purpose is to show that there exists $\tau^* = \lim_{n\to\infty}\tau^{(n)}$ and that τ^* is the parameterization of the invariant torus of the statement. According to proposition 3.2 and remark 3.3, we should verify that $\nabla e_1^* = e_2^* = e_3^* = 0$, where these error functions are defined as in equation (37), but now in terms of τ^* instead of τ. To apply proposition 3.2 to τ^*, we also need to verify that the invariance error e^* associated to τ^* (see (21)) is sufficiently small so that $\|(De^*)^{\top}D\tau_+^*(N_+^*)^{-1}\|_{\mathbb{T}^r} < 2$, where $N^* = (D\tau_+^*)^{\top}D\tau_+^*$. This latter condition is immediate from the proof.

Explicit expressions of one step of the iterative method. We consider the parameterization τ of a quasi-torus $\mathcal{T} = \tau(\mathbb{T}^r)$ of F, as well as the following associated objects: N, Ω, S, $\{e_j\}_{j=1}^{3}$, and e (see equations (6), (14), (17), and (18)). We suppose that N is an invertible matrix function along the torus and that $\det(\langle S\rangle_\theta) \neq 0$.

Although we are using the name τ for the parameterization, this may be not only the one of the statement of the theorem, but any element of the sequence $\{\tau^{(n)}\}_{n \geq 0}$. Then, we perform one step of the method to compute the new parameterization $\tau^{(1)}$ by definition 3.6. In particular, we are assuming that there are (small) solutions for a and b of the "generalized cohomological equations" in (41). We denote by $N^{(1)}$, $\Omega^{(1)}$, $S^{(1)}$, $\{e_j^{(1)}\}_{j=1}^3$, and $e^{(1)}$ the expressions mentioned above, but now associated to $\tau^{(1)}$ instead of τ. Further, we denote by $\bar{N}^{(1)}$, $\bar{\Omega}^{(1)}$, $\bar{S}^{(1)}$, $\{\bar{e}_j^{(1)}\}_{j=1}^3$, and $\bar{e}^{(1)}$ the corresponding objects associated to $\bar{\tau}$ (see (40)). At this stage of the proof, we are mainly concerned with the formal aspects of the method. We postpone the discussion of the quantitative conditions necessary to iterate until the next item of the proof. Our main purpose now is to provide formulas for $\{e_j^{(1)}\}_{j=1}^3$ that allow us to show that these errors behave quadratically with the step. To simplify the notations, in the upcoming formulas we follow the same criteria regarding to the dependence on θ and ω that we have explained in remark 3.1.

We begin by discussing several auxiliary aspects but that play a crucial role in the proof. Firstly, we relate the matrix Ω to e_2 through the difference equation (38) (we recall that $\langle \Omega \rangle_\theta = 0$):

$$\Omega = \frac{1}{2} \mathcal{L}_\omega^{-1}(De_2 - (De_2)^\top). \tag{75}$$

Formula (75) implies that if e_2 is small, then so is Ω. Next, we introduce the square matrix $\mathcal{N}$ defined as follows:

$$\mathcal{N} = (\mathrm{Id}_r + \Omega\, N^{-1}\, \Omega\, N^{-1})N. \tag{76}$$

Since the size of Ω is small, then we may suppose that $\mathcal{N}$ is an invertible matrix function along $\mathcal{T}$. The condition $\det(\mathcal{N}) \neq 0$ is equivalent to say that the columns of the matrices $D\tau$ and $JD\tau$ form a basis of $\mathbb{R}^{2r}$ at any point. Indeed, if Ω is small, then $JD\tau\, N^{-1}$ and $D\tau$ form an approximate symplectic basis. If $\xi = \xi(\theta)$ is a vector function, $\xi : \mathbb{T}^r \to \mathbb{R}^r$, then there are functions $\hat{\alpha}, \hat{\beta} : \mathbb{T}^r \to \mathbb{R}^r$ such that:

$$\xi = D\tau_+ \,\hat{\alpha} + JD\tau_+ \,\hat{\beta}. \tag{77}$$

If we multiply equation (77) by $N_+^{-1}(D\tau_+)^\top$ and $N_+^{-1}(D\tau_+)^\top J$, then we obtain the following expressions:

$$N_+^{-1}(D\tau_+)^\top \xi = \hat{\alpha} + N_+^{-1}\Omega_+ \,\hat{\beta}, \qquad N_+^{-1}(D\tau_+)^\top J\xi = N_+^{-1}\Omega_+ \,\hat{\alpha} - \hat{\beta}.$$

If $\det(\mathcal{N}) \neq 0$, then the unique solutions for $\hat{\alpha}$ and $\hat{\beta}$ of equation (77) are:

$$\hat{\alpha} = \mathcal{N}_+^{-1}((D\tau_+)^\top \xi + \Omega_+ N_+^{-1}(D\tau_+)^\top J\xi), \qquad \hat{\beta} = \mathcal{N}_+^{-1}(\Omega_+ N_+^{-1}(D\tau_+)^\top \xi - (D\tau_+)^\top J\xi). \tag{78}$$

We use computations above to relate the invariance error e with $\{e_j\}_{j=1}^3$. We express e as:

$$e = D\tau_+ \,\alpha + JD\tau_+ \,\beta, \tag{79}$$

for some vector functions α and β. Explicit expressions for α and β are provided by formulas in (78), just by replacing ξ by e. To relate α and β with $\{e_j\}_{j=1}^3$, we adapt the computations made in the proof of proposition 3.2 to the case of non-zero errors $\{e_j\}_{j=1}^3$. Explicitly, we consider the definitions of $\{e_j\}_{j=1}^3$ in (37) (recall that we also can write $e_1 = \tau_+^\top J e$) as well as (21) and (24), and we establish the following relations:

$$-(De)^\top J \tau_+ + (D\tau_+)^\top J e = \nabla e_1, \tag{80}$$
$$(De)^\top J \tau_+ + (D\tau_+)^\top J e = -e_2 - (De)^\top J e, \tag{81}$$
$$D\tau_+^\top e = e_3. \tag{82}$$

If we perform the sum of the formulas (80) and (81), then we obtain:

$$(D\tau_+)^\top J e = \frac{1}{2}\left(\nabla e_1 - e_2 - (De)^\top J e\right). \tag{83}$$

By replacing e (but not De) in equations (82) and (83) by the expression in (79), we obtain the following equations:

$$N_+\, \alpha + \Omega_+\, \beta = e_3, \qquad \hat{\Omega}_+\, \alpha - \hat{N}_+\, \beta = \frac{1}{2}\left(\nabla e_1 - e_2\right), \qquad (84)$$

where we introduce:

$$\hat{\Omega}_+ = \Omega_+ + \frac{1}{2}(De)^\top J\, D\tau_+, \qquad \hat{N}_+ = \left(\mathrm{Id}_r + \frac{1}{2}(De)^\top D\tau_+\, N_+^{-1}\right) N_+. \qquad (85)$$

At this point, we assume known an a priori estimate for the size of De, and that it is sufficiently small so that $\|(De)^\top\|_{\mathbb{T}^r} \cdot \|D\tau_+\, N_+^{-1}\|_{\mathbb{T}^r} < 2$. By the hypotheses of theorem 2.6, this inequality holds at least for the initial parameterization. Since we are going to show that the error e goes to zero almost quadratically with the step, this is indeed a very mild iterative assumption (i.e., a rough estimate of this size of De is sufficient). Consequently, since $\det(N) \neq 0$, we can also assume that $\det(\hat{N}) \neq 0$. Then, we introduce the following matrices:

$$\hat{\mathcal{N}}_1 = (\mathrm{Id}_r + \Omega\, \hat{N}^{-1}\, \hat{\Omega}\, N^{-1})\, N, \qquad \hat{\mathcal{N}}_2 = (\mathrm{Id}_r + \hat{\Omega}\, N^{-1}\, \Omega\, \hat{N}^{-1})\, \hat{N}, \qquad (86)$$

which are small perturbations of the matrix $\mathcal{N}$ defined in (76). Hence, we may also assume that $\det \hat{\mathcal{N}}_j \neq 0$, for $j = 1, 2$. We use these matrices and equations in (84) to establish the following formulas for α and β in (79):

$$\alpha = \hat{\mathcal{N}}_{1;+}^{-1}\left(e_3 + \frac{1}{2}\Omega_+\, \hat{N}_+^{-1}(\nabla e_1 - e_2)\right), \qquad \beta = \hat{\mathcal{N}}_{2;+}^{-1}\left(\hat{\Omega}_+\, N_+^{-1}\, e_3 - \frac{1}{2}(\nabla e_1 - e_2)\right). \qquad (87)$$

The last auxiliary computations of this first part of the proof refer to the automatic quasi-reducibility of τ. We use equation (24) to control the action of $DF(\tau)$ on the columns of $D\tau$ as $DF(\tau)D\tau = D\tau_+ + De$. To control the action of $DF(\tau)$ on the columns of $JD\tau\, N^{-1}$, we write:

$$DF(\tau)\, J\, D\tau\, N^{-1} = D\tau_+\, [S + A] + J\, D\tau_+\, \left[N_+^{-1} + B\right], \qquad (88)$$

for some square matrix functions A and B (that we expect are small), where S is the matrix defined in (14). To compute A and B, we use formulas in (77) and (78) with the following value of ξ:

$$\xi = DF(\tau)\, J\, D\tau\, N^{-1} - D\tau_+\, S - J\, D\tau_+\, N_+^{-1}.$$

We have that $(D\tau_+)^\top \xi = -\Omega_+\, N_+^{-1}$ and

$$\begin{aligned}
(D\tau_+)^\top J\xi &= (DF(\tau)\, D\tau - De)^\top J\, DF(\tau)\, J\, D\tau\, N^{-1} - \Omega_+\, S + \mathrm{Id}_r \\
&= -(De)^\top J\, DF(\tau)\, J\, D\tau\, N^{-1} - \Omega_+\, S,
\end{aligned}$$

where we use that $(DF)^\top J\, DF = J$. Then, we end up with the following expressions for A and B:

$$\begin{aligned}
A &= -\mathcal{N}_+^{-1}\Omega_+\, N_+^{-1}\left(\mathrm{Id}_r + (De)^\top J\, DF(\tau)\, J\, D\tau\, N^{-1} + \Omega_+\, S\right), \qquad &(89) \\
B &= \mathcal{N}_+^{-1}\left((De)^\top J DF(\tau) J D\tau N^{-1} + \Omega_+\, S - \Omega_+\, N_+^{-1}\Omega_+\, N_+^{-1}\right). \qquad &(90)
\end{aligned}$$

Next, we address the formulas of the errors $\{e_j^{(1)}\}_{j=1}^3$ associated to $\tau^{(1)}$. See definition 3.6 for the expression of $\tau^{(1)}$ and related objects. Firstly, we use Taylor's formula and equation (88) to expand $F(\bar{\tau}^{(1)})$ as follows:

$$F(\bar{\tau}^{(1)}) = F(\tau) + R_1 = F(\tau) + DF(\tau)\, \Delta\tau + R_2 = F(\tau) + D\tau_+\, S\, a + J D\tau_+\, N_+^{-1}\, a + \bar{R}_2, \qquad (91)$$

where, if we write $R_2 = (R_{2;1}, \ldots, R_{2;2r})$ and $F = (F_1, \ldots, F_{2r})$, then we have

$$R_1 = \left[\int_0^1 DF(\tau + s\Delta\tau)\, ds\right]\Delta\tau, \qquad \bar{R}_2 = R_2 + D\tau_+\, A\, a + JD\tau_+\, B\, a, \tag{92}$$

$$R_{2;j} = (\Delta\tau)^\top \left[\int_0^1 (1-s)D^2 F_j(\tau + s\Delta\tau)\, ds\right]\Delta\tau, \quad j = 1, \ldots, 2r. \tag{93}$$

We also consider the following expressions for $F(\tau^{(1)}) = F(\bar{\tau}^{(1)}(\theta + b))$ (recall that $F(\tau) = \tau_+ + e$):

$$\begin{aligned}
F(\tau^{(1)}) &= F(\tau(\theta + b)) + R_1(\theta + b) = \tau(\theta + \omega + b) + e(\theta + b) + R_1(\theta + b) \\
&= F(\tau(\theta + b)) + \big(D\tau(\theta + \omega + b)S(\theta + b) + JD\tau(\theta + \omega + b)N^{-1}(\theta + \omega + b)\big)\, a(\theta + b) \\
&\quad + \bar{R}_2(\theta + b).
\end{aligned}$$

By using (91), we obtain (we note that since $\bar{e}_1^{(1)}$ is a scalar valued function it does not change if we transpose it):

$$\bar{e}_1^{(1)} = (\bar{\tau}_+^{(1)})^\top J\, F(\bar{\tau}^{(1)}) = (\tau_+ + JD\tau_+ N_+^{-1} a_+)^\top J\, F(\bar{\tau}^{(1)}) = e_1 + \tau_+^\top \big(JD\tau_+ Sa + D\tau_+ N_+^{-1}\mathcal{L}_\omega a\big) + \hat{e}_1, \tag{94}$$

where

$$\hat{e}_1 = \tau_+^\top J\bar{R}_2 + a_+^\top N_+^{-1} D\tau_+^\top (e + R_1). \tag{95}$$

We relate $e_1^{(1)}$ to $\bar{e}_1^{(1)}$ through the following expressions:

$$\begin{aligned}
e_1^{(1)} &= (\tau_+^{(1)})^\top J\, F(\tau^{(1)}) = (\bar{\tau}^{(1)}(\theta + \omega + b_+))^\top J\, F(\bar{\tau}^{(1)}(\theta + b)) \\
&= \bar{e}_1^{(1)}(\theta + b) + (\bar{\tau}^{(1)}(\theta + \omega + b_+) - \bar{\tau}^{(1)}(\theta + \omega + b))^\top J\, F(\tau^{(1)}) \\
&= \bar{e}_1^{(1)}(\theta + b) - (\tau(\theta + \omega + b))^\top J\, D\tau(\theta + \omega + b)\,\mathcal{L}_\omega b + \hat{e}_4,
\end{aligned} \tag{96}$$

where we have written

$$\bar{\tau}^{(1)}(\theta + \omega + b_+) - \bar{\tau}^{(1)}(\theta + \omega + b) = D\tau(\theta + \omega + b)\,\mathcal{L}_\omega b + \hat{e}_2 + \hat{e}_3,$$

and

$$\begin{aligned}
\hat{e}_{2;j} &= \tau_j(\theta + \omega + b_+) - \tau_j(\theta + \omega + b) - D\tau_j(\theta + \omega + b)\,\mathcal{L}_\omega b \\
&= (\mathcal{L}_\omega b)^\top \left(\int_0^1 (1-s)D^2\tau_j(\theta + \omega + sb_+ + (1-s)b)\, ds\right)\mathcal{L}_\omega b, \qquad j = 1, \ldots, 2r, \tag{97}
\end{aligned}$$

$$\hat{e}_3 = \Delta\tau(\theta + \omega + b_+) - \Delta\tau(\theta + \omega + b) = \left(\int_0^1 D(\Delta\tau)(\theta + \omega + sb_+ + (1-s)b)\, ds\right)\mathcal{L}_\omega b, \tag{98}$$

$$\hat{e}_4 = (\mathcal{L}_\omega b)^\top (D\tau(\theta + \omega + b))^\top J\, (e(\theta + b) + R_1(\theta + b)) + (\hat{e}_2 + \hat{e}_3)^\top JF(\tau^{(1)}). \tag{99}$$

We apply the expression in (94) and equations for a and b in (41) to formula (96). We obtain (recall $e_1 = \tau_+^\top J\, e$):

$$e_1^{(1)} = (\tau(\theta + \omega + b))^\top J\, \hat{e}_5(\theta + b) + \hat{e}_1(\theta + b) + \hat{e}_4, \tag{100}$$

where

$$\hat{e}_5 = e - JD\tau_+ N_+^{-1}\left(\frac{1}{2}e_2 - \frac{1}{2}\nabla e_1\right) - D\tau_+ N_+^{-1} e_3. \tag{101}$$

To show that $\hat{e}_5$ has quadratic size we rely, once again, on formulas in (77) and (78), now taking $\xi = \hat{e}_5$. We write:

$$\hat{e}_5 = D\tau_+\, \hat{\alpha}_5 + JD\tau_+\, \hat{\beta}_5, \tag{102}$$

and we compute (see (83) and recall that $e_3 = (D\tau_+)^\top e$):

$$(D\tau_+)^\top \hat{e}_5 = -\Omega_+ N_+^{-1} \left(\frac{1}{2} e_2 - \frac{1}{2} \nabla e_1 \right), \qquad (D\tau_+)^\top J \hat{e}_5 = - \left(\frac{1}{2} (De)^\top Je + \Omega_+ N_+^{-1} e_3 \right).$$

Then, we obtain:

$$\hat{\alpha}_5 = -\mathcal{N}_+^{-1} \Omega_+ N_+^{-1} \left(\frac{1}{2} \left(e_2 - \nabla e_1 + (De)^\top Je \right) + \Omega_+ N_+^{-1} e_3 \right), \tag{103}$$

$$\hat{\beta}_5 = \mathcal{N}_+^{-1} \left(-\frac{1}{2} \Omega_+ N_+^{-1} \Omega_+ N_+^{-1} (e_2 - \nabla e_1) + \Omega_+ N_+^{-1} e_3 + \frac{1}{2} (De)^\top Je \right). \tag{104}$$

Remark 4.1. *Equations for a and b in (41) were introduced to assure that $e_2^{(1)}$ and $e_3^{(1)}$ have quadratic size. The fact that $e_1^{(1)}$ turns out to be also quadratic is due to the geometric constraints of exact symplectic maps. Furthermore, the fact that equation for b in (41) is a nonlinear equation plays a crucial role to guarantee that in equation (100) the contribution of b appears through a composition. The same comment applies to $e_2^{(1)}$ and $e_3^{(1)}$.*

Next, we focus on $\bar{e}_2^{(1)}$ (see (37)):

$$\bar{e}_2^{(1)} = (D\bar{\tau}_+^{(1)})^\top J \bar{\tau}_+^{(1)} - (D\bar{\tau}^{(1)})^\top (DF(\bar{\tau}^{(1)}))^\top J F(\bar{\tau}^{(1)}). \tag{105}$$

Since dealing with compact expressions seems a good way to optimize the computations, we consider the following expression for the second term in (105):

$$(D\bar{\tau}^{(1)})^\top (DF(\bar{\tau}^{(1)}))^\top J F(\bar{\tau}^{(1)}) = (D(F(\bar{\tau}^{(1)})))^\top J F(\bar{\tau}^{(1)}).$$

Besides, to simplify the upcoming computations, we also introduce some auxiliary notations:

$$\bar{a} = N^{-1} a, \qquad \hat{a} = N_+^{-1} a, \qquad c = Sa, \qquad \hat{c} = N_+^{-1} e_3. \tag{106}$$

We use definitions in equation (106) to establish the following formulas (see equation (91)):

$$\Delta\tau = J D\tau\, \bar{a} = J \sum_{j=1}^{r} \bar{a}_j \partial_{\theta_j} \tau, \qquad D(\Delta\tau) = J D\tau\, D\bar{a} + J \sum_{j=1}^{r} \bar{a}_j \partial_{\theta_j}(D\tau), \tag{107}$$

$$F(\bar{\tau}^{(1)}) = F(\tau) + D\tau_+\, c + J D\tau_+ \hat{a} + \bar{R}_2 = F(\tau) + \sum_{j=1}^{r} c_j \partial_{\theta_j} \tau_+ + J \sum_{j=1}^{r} \hat{a}_j \partial_{\theta_j} \tau_+ + \bar{R}_2, \tag{108}$$

$$D(F(\bar{\tau}^{(1)})) = DF(\tau)\, D\tau + D\tau_+\, Dc + \sum_{j=1}^{r} c_j \partial_{\theta_j}(D\tau_+) + J D\tau_+\, D\hat{a} + J \sum_{j=1}^{r} \hat{a}_j \partial_{\theta_j}(D\tau_+) + D\bar{R}_2. \tag{109}$$

By using formulas in (107), we obtain:

$$\begin{aligned}
(D\bar{\tau}^{(1)})^\top J \bar{\tau}^{(1)} &= (D\tau + D(\Delta\tau))^\top J (\tau + \Delta\tau) \\
&= (D\tau)^\top J\tau - a + (D\bar{a})^\top (D\tau)^\top \tau + \sum_{j=1}^{r} \bar{a}_j \partial_{\theta_j}(D\tau)^\top \tau + \hat{e}_6,
\end{aligned} \tag{110}$$

where

$$\hat{e}_6 = (D(\Delta\tau))^\top J \Delta\tau. \tag{111}$$

We also compute (see equations (21), (24), (91), (106), (108), and (109)):

$$
\begin{aligned}
(D(F(\bar{\tau}^{(1)})))^\top J\, F(\bar{\tau}^{(1)}) &= (DF(\tau)\, D\tau + DR_1)^\top J\, (F(\tau) + R_1) \\
&= (D\tau)^\top (DF(\tau))^\top J\, F(\tau) + (DF(\tau)\, D\tau)^\top J\, R_1 + (DR_1)^\top J F(\tau) + (DR_1)^\top J R_1 \\
&= (D\tau)^\top (DF(\tau))^\top J\, F(\tau) - a + (Dc)^\top (D\tau_+)^\top J\, \tau_+ + (D\hat{a})^\top (D\tau_+)^\top \tau_+ \\
&\quad + \sum_{j=1}^{r} c_j \partial_{\theta_j} (D\tau_+)^\top J\, \tau_+ + \sum_{j=1}^{r} \hat{a}_j \partial_{\theta_j} (D\tau_+)^\top \tau_+ + \hat{e}_7,
\end{aligned}
\tag{112}
$$

where

$$
\hat{e}_7 = \Omega_+\, c + (D\tau_+)^\top J\, \bar{R}_2 + (De)^\top J\, R_1 + (D\bar{R}_2)^\top J\tau_+ + (DR_1)^\top J\, e + (DR_1)^\top J\, R_1.
\tag{113}
$$

If we apply expressions (110) and (112) to (105), then we obtain:

$$
\begin{aligned}
\bar{e}_2^{(1)} &= e_2 - \mathcal{L}_\omega a + \sum_{j=1}^{r} (\bar{a}_{+;j} - \hat{a}_j)\, \partial_{\theta_j} (D\tau_+)^\top \tau_+ + (D\bar{a}_+ - D\hat{a})^\top (D\tau_+)^\top \tau_+ \\
&\quad - (Dc)^\top (D\tau_+)^\top J\, \tau_+ - \sum_{j=1}^{r} c_j \partial_{\theta_j} (D\tau_+)^\top J\, \tau_+ + \hat{e}_{6;+} - \hat{e}_7.
\end{aligned}
\tag{114}
$$

In order to express $e_2^{(1)}$ in terms of $\bar{e}_2^{(1)}$, a crucial observation is that:

$$
D\tau^{(1)} = D\bar{\tau}^{(1)}(\theta + b)\, (\mathrm{Id}_r + Db).
\tag{115}
$$

We obtain:

$$
\begin{aligned}
e_2^{(1)} &= (D\tau_+^{(1)})^\top J\, \tau_+^{(1)} - (D\tau^{(1)})^\top (DF(\tau^{(1)}))^\top J\, F(\tau^{(1)}) \\
&= (\mathrm{Id}_r + Db_+)^\top (D\bar{\tau}^{(1)}(\theta + \omega + b_+))^\top J\, \bar{\tau}^{(1)}(\theta + \omega + b_+) \\
&\quad - (\mathrm{Id}_r + Db)^\top (D\bar{\tau}^{(1)}(\theta + b))^\top DF(\bar{\tau}^{(1)}(\theta + b))^\top J\, F(\bar{\tau}^{(1)}(\theta + b)) \\
&= (\mathrm{Id}_r + Db)^\top \left(\bar{e}_2^{(1)}(\theta + b) + D\left((D\bar{\tau}^{(1)})^\top J\, \bar{\tau}^{(1)}\right)(\theta + \omega + b)\, \mathcal{L}_\omega b + \hat{R}_2 \right) \\
&\quad + (\mathcal{L}_\omega (Db))^\top \left((D\bar{\tau}^{(1)}(\theta + \omega + b))^\top J\, \bar{\tau}^{(1)}(\theta + \omega + b) + \bar{R}_1 \right).
\end{aligned}
\tag{116}
$$

where

$$
\begin{aligned}
\bar{R}_1 &= (D\bar{\tau}^{(1)}(\theta + \omega + b_+))^\top J\, \bar{\tau}^{(1)}(\theta + \omega + b_+) - (D\bar{\tau}^{(1)}(\theta + \omega + b))^\top J\, \bar{\tau}^{(1)}(\theta + \omega + b), \\
\hat{R}_2 &= (D\bar{\tau}^{(1)}(\theta + \omega + b_+))^\top J\, \bar{\tau}^{(1)}(\theta + \omega + b_+) - (D\bar{\tau}^{(1)}(\theta + \omega + b))^\top J\, \bar{\tau}^{(1)}(\theta + \omega + b) \\
&\quad - D\left((D\bar{\tau}^{(1)})^\top J\, \bar{\tau}^{(1)}\right)(\theta + \omega + b)\, \mathcal{L}_\omega b.
\end{aligned}
$$

To control $\bar{R}_1$ and $\hat{R}_2$, we are going to use the following expressions for their components:

$$
\bar{R}_{1;j} = \left[\int_0^1 D\left(\langle \partial_{\theta_j} \bar{\tau}^{(1)}, J\, \bar{\tau}^{(1)} \rangle\right)(\theta + \omega + s b_+ + (1-s)b)\, ds \right] \mathcal{L}_\omega b,
\tag{117}
$$

$$
\hat{R}_{2;j} = (\mathcal{L}_\omega b)^\top \left[\int_0^1 (1-s) D^2 \left(\langle \partial_{\theta_j} \bar{\tau}^{(1)}, J\, \bar{\tau}^{(1)} \rangle\right)(\theta + \omega + s b_+ + (1-s)b)\, ds \right] \mathcal{L}_\omega b.
\tag{118}
$$

To simplify formula (116) for $e_2^{(1)}$, we observe that from the equation of b in (41) we derive the following expressions (see definitions in (106)):

$$\mathcal{L}_\omega b = c(\theta + b) + \hat{c}(\theta + b), \qquad \mathcal{L}_\omega(Db) = (Dc(\theta + b) + D\hat{c}(\theta + b))\,(\mathrm{Id}_r + Db). \tag{119}$$

Consequently, we obtain:

$$e_2^{(1)} = (\mathrm{Id} + Db)^\top\,(\hat{e}_8(\theta + b) + \hat{e}_9)\,, \tag{120}$$

where:

$$\hat{e}_8 = \bar{e}_2^{(1)} + \left(D\left((D\bar{\tau}_+^{(1)})^\top J\,\bar{\tau}_+^{(1)}\right)\right)(c + \hat{c}) + (Dc + D\hat{c})^\top (D\bar{\tau}_+^{(1)})^\top J\,\bar{\tau}_+^{(1)}, \tag{121}$$

$$\hat{e}_9 = (Dc(\theta + b) + D\hat{c}(\theta + b))^\top \bar{R}_1 + \hat{R}_2. \tag{122}$$

By definition, we have that $\hat{e}_9$ is a quadratic error. We have to show that $\hat{e}_8$ is quadratic too. We rewrite expression (121) for $\hat{e}_8$ in the following form (see equation (114)):

$$\hat{e}_8 = e_2 - \mathcal{L}_\omega a + \sum_{j=1}^r (\bar{a}_{+;j} - \hat{a}_j)\,\partial_{\theta_j}(D\tau_+)^\top \tau_+ + (D\bar{a}_+ - D\hat{a})^\top (D\tau_+)^\top \tau_+$$

$$- \sum_{j=1}^r c_j \partial_{\theta_j}(D\tau_+)^\top J\,\tau_+ + \left(D\left((D\tau_+)^\top J\,\tau_+\right)\right)(c + \hat{c}) + (D\hat{c})^\top (D\tau_+)^\top J\,\tau_+ + \hat{e}_{10}, \tag{123}$$

where

$$\hat{e}_{10} = \hat{e}_{6;+} - \hat{e}_7 + \left(D\left((D\bar{\tau}_+^{(1)})^\top J\,\bar{\tau}_+^{(1)} - (D\tau_+)^\top J\,\tau_+\right)\right)(c + \hat{c})$$

$$+ (Dc + D\hat{c})^\top \left((D\bar{\tau}_+^{(1)})^\top J\,\bar{\tau}_+^{(1)} - (D\tau_+)^\top J\,\tau_+\right). \tag{124}$$

To simplify formula (123) for $\hat{e}_8$, and to make it clear that it is a quadratic error, we compute the gradient of the following two scalar expressions. Firstly:

$$\nabla\left[\tau_+^\top D\tau_+\,(\bar{a}_+ - \hat{a})\right] = (D\bar{a}_+ - D\hat{a})^\top (D\tau_+)^\top \tau_+ + \sum_{j=1}^r (\bar{a}_{+;j} - \hat{a}_j)\,\partial_{\theta_j}(D\tau_+)^\top \tau_+ + \mathcal{L}_\omega\,a,$$

where we have used definitions in (106) to rewrite $\bar{a}_+ - \hat{a} = N_+^{-1}\,\mathcal{L}_\omega a$. Secondly:

$$\nabla\left[\tau_+^\top J\,D\tau_+\,\hat{c}\right] = -(D\hat{c})^\top (D\tau_+)^\top J\,\tau_+ - \left(D\left((D\tau_+)^\top J\,\tau_+\right)\right)^\top \hat{c}.$$

Moreover, we also have:

$$\left(D\left((D\tau_+)^\top J\,\tau_+\right)\right)c = \sum_{j=1}^r c_j \partial_{\theta_j}(D\tau_+)^\top J\,\tau_+ + \Omega_+\,c.$$

Using computations above, we obtain the following expression:

$$\hat{e}_8 = e_2 - 2\mathcal{L}_\omega a + \nabla\left[\tau_+^\top D\tau_+\,(\bar{a}_+ - \hat{a}) - \tau_+^\top J\,D\tau_+\,\hat{c}\right] + \left(D\vartheta_+ - (D\vartheta_+)^\top\right)\hat{c} + \Omega_+\,c + \hat{e}_{10},$$

where we recall that $\vartheta = (D\tau)^\top J\,\tau$ (see equation (6)). Using the relation $2\,\Omega = D\vartheta - (D\vartheta)^\top$ (see definition 2.2), formulas introduced in equation (106), the equation for a in (41), and that $e_1 = \tau_+^\top J\,e$, we conclude:

$$\hat{e}_8 = \nabla\left[\tau_+^\top J\,\hat{e}_5\right] + 2\Omega_+\,N_+^{-1}\,e_3 + \Omega_+\,S\,a + \hat{e}_{10}, \tag{125}$$

where $\hat{e}_5$ (for which we have already established that it has quadratic size) is defined in equation (101).

To discuss $e_3^{(1)}$, we introduce (see (21), (40) and (91)):

$$\bar{e}^{(1)} = F(\bar{\tau}^{(1)}) - \bar{\tau}_+^{(1)} = e + R_1 - \Delta\tau_+, \tag{126}$$

$$\bar{e}_3^{(1)} = (D\bar{\tau}_+^{(1)})^\top \left(F(\bar{\tau}^{(1)}) - \bar{\tau}_+^{(1)}\right) = e_3 + N_+ \, S \, a + \hat{e}_{11}, \tag{127}$$

where

$$\hat{e}_{11} = -\Omega_+ \, N_+^{-1} \, \mathcal{L}_\omega \, a + (D\tau_+)^\top \, \bar{R}_2 + (D(\Delta\tau_+))^\top \, \bar{e}^{(1)}. \tag{128}$$

Additionally, we also introduce:

$$\hat{e}_{12} = \bar{\tau}^{(1)}(\theta + \omega + b_+) - \bar{\tau}^{(1)}(\theta + \omega + b) - D\bar{\tau}^{(1)}(\theta + \omega + b) \, \mathcal{L}_\omega \, b,$$
$$\hat{e}_{13} = D\bar{\tau}^{(1)}(\theta + \omega + b_+) - D\bar{\tau}^{(1)}(\theta + \omega + b), \tag{129}$$

whose components can be expressed as follows:

$$\hat{e}_{12;j} = (\mathcal{L}_\omega b)^\top \left[\int_0^1 (1-s)D^2\bar{\tau}_j^{(1)}(\theta + \omega + sb_+ + (1-s)b) \, ds\right] \mathcal{L}_\omega b, \tag{130}$$

$$\hat{e}_{13;j,l} = \left[\int_0^1 D\left(\partial_{\theta_l}\bar{\tau}_j^{(1)}\right)(\theta + \omega + sb_+ + (1-s)b) \, ds\right] \mathcal{L}_\omega b. \tag{131}$$

Then, we can write $e_3^{(1)}$ as follows (see (115)):

$$\begin{aligned} e_3^{(1)} &= (D\tau_+^{(1)})^\top \left(F(\tau^{(1)}) - \tau_+^{(1)}\right) \\ &= (\mathrm{Id} + Db_+)^\top (D\bar{\tau}^{(1)}(\theta + \omega + b_+))^\top \left(\bar{e}^{(1)}(\theta + b) - D\bar{\tau}^{(1)}(\theta + \omega + b) \, \mathcal{L}_\omega \, b - \hat{e}_{12}\right) \\ &= (\mathrm{Id} + Db_+)^\top \left(\hat{e}_{14}(\theta + b) + \hat{e}_{15}\right), \end{aligned} \tag{132}$$

where (see (106), (119), (127), and (129)):

$$\hat{e}_{14} = \bar{e}_3^{(1)} - \bar{N}_+^{(1)}(c + \hat{c}) = \hat{e}_{11} - (\bar{N}_+^{(1)} - N_+)(S \, a + N_+^{-1} \, e_3) = \hat{e}_{11} - \Delta\bar{N}_+ (S \, a + N_+^{-1} \, e_3), \tag{133}$$

$$\hat{e}_{15} = -(D\bar{\tau}^{(1)}(\theta + \omega + b_+))^\top \hat{e}_{12} + (\hat{e}_{13})^\top \left(\bar{e}^{(1)}(\theta + b) - D\bar{\tau}^{(1)}(\theta + \omega + b) \, \mathcal{L}_\omega \, b\right), \tag{134}$$

and where we are also introducing:

$$\bar{N}^{(1)} = (D\bar{\tau}^{(1)})^\top D\bar{\tau}^{(1)} = N + \Delta\bar{N} = N(\mathrm{Id}_r + N^{-1}\Delta\bar{N}), \tag{135}$$

$$\Delta\bar{N} = (D\tau)^\top D(\Delta\tau) + (D(\Delta\tau))^\top D(\Delta\tau) + (D(\Delta\tau))^\top D(\Delta\tau). \tag{136}$$

Bounding one step of the iterative method. Let τ be the parameterization of the quasi-torus of the statement of theorem 2.6 (so bounds from (11) to (19) are met for it). We now carry out the quantitative estimates on expressions above for $\{e_j^{(1)}\}_{j=1}^3$, after applying one step of the method to τ. Our main purpose is to construct a constant m for which theorem 2.6 holds. To do that, we proceed analogously as we have done to construct $\bar{m}$ in the proof of proposition 3.7. Hence, the value of m is redefined recursively to meet a finite number of conditions and the last value obtained is the one of the statement. Since all the parameterization of the sequence $\{\tau^{(n)}\}_{n\geq 0}$ will verify analogous bounds as those of the initial one, with the same constants $\{m_j\}_{j=1}^9$, we are going to see that the value of m thus obtained is not only valid for τ, but that can be used iteratively. Indeed, the quantities that change iteratively with n are the values of ρ and $\{\mu_j\}_{j=1}^3$ (but keeping the conditions $\rho^{(n)} > 0$ and $0 \leq \mu_j^{(n)} \leq 1$). To

perform the forthcoming bounds, we use proposition 2.5 to control the action of the small-divisors operator $\mathcal{L}_\omega^{-1}$, as well as some basic properties of the norm $\|\cdot\|_\rho$ (see definition 2.4 and equations (51) and (52)). Besides, we also use proposition 3.7 to control the solutions for $\langle a \rangle_\theta$ and b of the generalized cohomological equations in (41).

Firstly, we use the constants on the statement of theorem 2.6 to introduce the normalized error

$$\mu = \mu_3 + \frac{\mu_1 + \mu_2}{\gamma \delta^\nu}. \tag{137}$$

Suppose that the following bounds hold for all the elements of the finite sequence of values of m we are going to construct:

$$m \frac{\mu}{\gamma \delta^{\nu+1}} < \min\left\{\frac{1}{2}, \frac{m_8}{m_5 \cdot m_6}, \Theta\right\}, \tag{138}$$

where Θ is defined in (16) (we also recall that $\gamma < 1$ and $\delta < 1$). Then, we apply proposition 2.5 and the Cauchy estimates to formula (75). We obtain:

$$\|\Omega\|_{\rho-2\delta} = \frac{1}{2}\|\mathcal{L}_\omega^{-1}(De_2 - (De_2)^\top)\|_{\rho-2\delta} \leq \frac{\sigma}{2}\frac{\|De_2 - (De_2)^\top\|_{\rho-\delta}}{\gamma \delta^\nu} \leq \frac{r\sigma\|e_2\|_\rho}{\gamma \delta^{\nu+1}} \leq \frac{r\sigma\mu_2}{\gamma \delta^{\nu+1}} \leq m\frac{\mu}{\delta}.$$

By redefining m, we also have that $\|\Omega N^{-1}\|_{\rho-2\delta} \leq m\mu/\delta$. Since condition (138) implies that $\|\Omega N^{-1}\|_{\rho-2\delta} \leq 1/\sqrt{2}$, by Neumann's series we have that the matrix $\mathcal{N}$ of (76) is non-singular along the torus and verifies:

$$\|\mathcal{N}^{-1}\|_{\rho-2\delta} \leq \frac{\|N^{-1}\|_\rho}{1 - \|\Omega N^{-1}\|^2_{\rho-2\delta}} \leq 2m_6.$$

We consider formulas in (85). By the hypotheses of theorem 2.6, we have:

$$\max\{\|(De)^\top D\tau_+ N_+^{-1}\|_\rho, \|(De)^\top J D\tau_+ N_+^{-1}\|_\rho\} \leq \|(De)^\top\|_\rho \cdot m_5 \cdot m_6 \leq m_8 < 2. \tag{139}$$

By Neumann's series we conclude that $\hat{N}$ is non-singular along the torus and that $\|\hat{N}^{-1}\|_\rho \leq 2m_6$. Morerover, we also obtain that $\|\hat{\Omega} N^{-1}\|_{\rho-2\delta} \leq 2$. We proceed analogously for formulas in (86). We have:

$$\max\{\|\Omega \hat{N}^{-1} \hat{\Omega} N^{-1}\|_{\rho-2\delta}, \|\hat{\Omega} N^{-1} \Omega \hat{N}^{-1}\|_{\rho-2\delta}\} \leq m\frac{\mu}{\delta}.$$

Using that $m\mu/\delta \leq 1/2$, we conclude that $\|\hat{\mathcal{N}}_j^{-1}\|_{\rho-2\delta} \leq m$, $j = 1, 2$. Consequently, we can use expressions in (79) and (87) to control the size of the invariance error e and to derive a refined bound for De:

$$\|\alpha\|_{\rho-2\delta} \leq m\mu, \quad \|\beta\|_{\rho-2\delta} \leq m\mu, \quad \|e\|_{\rho-2\delta} \leq m\mu, \quad \|De\|_{\rho-3\delta} \leq m\frac{\mu}{\delta}.$$

Next, we focus on the (automatic) quasi-reducibility of τ. We have (see (14), (89), and (90), and use that $\mu/\delta \leq 1$):

$$\|S\|_\rho \leq m, \quad \|A\|_{\rho-3\delta} \leq m\frac{\mu}{\delta}, \quad \|B\|_{\rho-3\delta} \leq m\frac{\mu}{\delta}.$$

Now, we address the generalized cohomological equations for a and b introduced in definition 3.6. Firstly, by applying proposition 2.5 to the first equation in (41), we obtain $\|\tilde{a}\|_{\rho-\delta} \leq m\mu$. Next, to solve for b and $\xi = \langle a \rangle_\theta$ the second equation in (41), we apply proposition 3.7. To do that, we take $\hat{e} = N_+^{-1} e_3$, the same functions $\tilde{a}$ and S that we are considering, $\bar\rho = \rho - 3\delta > 0$, the same frequency vector ω, and the same values of δ, γ, ν, and μ that we are considering. Conditions $m \geq 1$, $0 < \delta < \min\{1, \bar\rho/4\}$, and $0 < \mu < 1$ are all straightforward. Then, once again redefining the value of m, the bounds in (46) and (47) are immediate. Finaly, condition (48) follows from hypothesis (138). Hence, equation (50) provides the following bounds (for an appropriate value of m):

$$|\langle a \rangle_\theta| \leq m\mu, \quad \|a\|_{\rho-\delta} \leq m\mu, \quad \|b\|_{\rho-7\delta} \leq m\frac{\mu}{\gamma \delta^\nu}, \quad \|Db\|_{\rho-7\delta} \leq m\frac{\mu}{\gamma \delta^{\nu+1}}.$$

Moreover, we also have (see equation (40)):

$$\|\Delta\tau\|_{\rho-\delta} \leq m\mu, \quad \|D(\Delta\tau)\|_{\rho-2\delta} \leq m\frac{\mu}{\delta}.$$

To guarantee that the new parameterization $\tau^{(1)}$ (see equations (39) and (40)) is well defined and to control other compositions involved in the forthcoming expressions, we observe that hypothesis (138) means that $\|b\|_{\rho-7\delta} \leq \delta$. Hence, we have:

$$(\cdot + f_s(\cdot))(\Delta(\rho - 7\delta)) \subset \Delta(\rho - 6\delta), \qquad \forall s \in [0,1], \tag{140}$$

where $f_s(\theta)$ can be either $f_s = \theta + sb$ or $f_s = \theta + \omega + sb_+ + (1-s)b$. Inclusions in (140) also hold in the closures of $\Delta(\rho - 7\delta)$ and $\Delta(\rho - 6\delta)$. Since we have that $\bar{\tau}^{(1)} = \tau + \Delta\tau$ is well-defined in the closure of $\Delta(\rho - \delta)$, we conclude that $\tau^{(1)}(\theta) = \bar{\tau}^{(1)}(\theta + b(\theta))$ is analytic in $\Delta(\rho - 7\delta)$. To control $\tau^{(1)} - \tau$, we write it as follows:

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

We obtain the bounds:

$$\|\tau^{(1)} - \tau\|_{\rho-7\delta} \leq m\frac{\mu}{\gamma\delta^\nu}, \qquad \|D\tau^{(1)} - D\tau\|_{\rho-8\delta} \leq m\frac{\mu}{\gamma\delta^{\nu+1}}.$$

Since condition (138) implies that $m\mu/(\gamma\delta^\nu) \leq m_4 - \|\tau\|_\rho$ and $m\mu/(\gamma\delta^{\nu+1}) < m_5 - \|D\tau\|_\rho$, we also obtain:

$$\|\bar{\tau}^{(1)}\|_{\rho-\delta} < m_4, \quad \|D\bar{\tau}^{(1)}\|_{\rho-2\delta} < m_5, \quad \|\tau^{(1)}\|_{\rho-7\delta} < m_4, \quad \|D\tau^{(1)}\|_{\rho-8\delta} < m_5.$$

Condition (138) also means that $\max\{\|\Delta\tau\|_{\rho-\delta}, \|\tau^{(1)} - \tau\|_{\rho-7\delta}\} \leq m\mu/(\gamma\delta^\nu) \leq d_{\tau,\rho,\mathcal{U}} - m_7$, which implies that

$$d_{\tau^{(1)},\rho-7\delta,\mathcal{U}} > m_7, \quad d_{\tau+s\Delta\tau,\rho-\delta,\mathcal{U}} > m_7, \quad d_{\tau+s(\tau^{(1)}-\tau),\rho-7\delta,\mathcal{U}} > m_7, \quad \forall s \in [0,1],$$

where we are using definition (13). These bounds mean that $\tau^{(1)}(\Delta(\rho - 7\delta)) \subset \mathcal{U}$, $(\tau + s\Delta\tau)(\Delta(\rho - \delta)) \subset \mathcal{U}$, and that $(\tau + s(\tau^{(1)} - \tau))(\Delta(\rho - 7\delta)) \subset \mathcal{U}$, $\forall s \in [0,1]$. Hence, we can use that $\|F\|_{\mathcal{U}} \leq m_1$ (see (11)) to control $\|F(\tau^{(1)})\|_{\rho-7\delta} \leq m_1$, $\|F(\tau + s\Delta\tau)\|_{\rho-\delta} \leq m_1$, and $\|F(\tau + s(\tau^{(1)} - \tau))\|_{\rho-7\delta} \leq m_1$, as well as the analogous expressions for DF and D^2F_j that follow from the other inequalities in (11).

To ensure that we can iterate definition 3.6, we should also control the invariance error of $\tau^{(1)}$. We do not use $e^{(1)}$ as an indicator of how far $\mathcal{T}^{(1)} = \tau^{(1)}(\mathbb{T}^r)$ is from being an invariant torus (this role is played by $\{e_j^{(1)}\}_{j=1}^3$), but we need to guarantee that the first condition in (15) still holds if we replace e by $e^{(1)}$ and ρ by $\rho - 8\delta$ (e.g., this condition for e has already been used in (139)). To do that, computing a rough estimate for $\|(De^{(1)})^\top\|_{\rho^{(1)}}$ is sufficient. We consider the following expressions (see (39), (40), and (126), and compare with (129) and (131)):

$$e^{(1)} = F(\tau^{(1)}) - \tau_+^{(1)} = F(\bar{\tau}^{(1)}(\theta + b)) - \bar{\tau}^{(1)}(\theta + \omega + b_+) = \bar{e}^{(1)}(\theta + b) - \hat{e}_{16},$$

$$\hat{e}_{16} = \bar{\tau}^{(1)}(\theta + \omega + b_+) - \bar{\tau}^{(1)}(\theta + \omega + b) = \left[\int_0^1 D\bar{\tau}^{(1)}(\theta + \omega + sb_+ + (1-s)b)\,ds\right]\mathcal{L}_\omega b.$$

Then, we consider the following estimates (here we are using the refined bound $\|e\|_{\rho-2\delta} \leq m\mu$):

$$\|R_1\|_{\rho-\delta} \leq m\mu, \quad \|\bar{e}^{(1)}\|_{\rho-2\delta} \leq m\mu, \quad \|\mathcal{L}_\omega b\|_{\rho-7\delta} \leq m\mu, \quad \|\hat{e}_{16}\|_{\rho-7\delta} \leq m\mu, \quad \|e^{(1)}\|_{\rho-7\delta} \leq m\mu,$$

where to bound $\mathcal{L}_\omega b$ we are using the expression for it provided by the second equation in (41). Consequently, we obtain the (rough) estimate $\|(De^{(1)})^\top\|_{\rho-8\delta} \leq m\mu/\delta$. Since condition (138) implies that $m\mu/\delta \leq m_8/(m_5 \cdot m_6)$, the desired estimate $\|(De^{(1)})^\top\|_{\rho-8\delta} \cdot m_5 \cdot m_6 \leq m_8$ is guaranteed.

Next, we show that $N^{(1)} = (D\tau^{(1)})^\top D\tau^{(1)}$ is non-singular along the tous and that the size of its inverse, in the set $\Delta(\rho^{(1)} - 7\delta)$, is still controlled by m_6. Using (115) and (135), we write:

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

Then, we derive the following bounds (see (136)), which include a rough estimate for the size of $(N^{(1)})^{-1}$:

$$\|\Delta\bar{N}\|_{\rho-2\delta} \le m\frac{\mu}{\delta}, \quad \|N^{-1}\Delta\bar{N}\|_{\rho-2\delta} \le m\frac{\mu}{\delta}, \quad \|\bar{N}^{(1)}\|_{\rho-2\delta} \le m, \quad \|(\mathrm{Id}_r + N^{-1}\Delta\bar{N})^{-1}\|_{\rho-2\delta} \le 2,$$

$$\|(\mathrm{Id}_r + Db)^{-1}\|_{\rho-7\delta} \le 2, \quad \|(\bar{N}^{(1)})^{-1}\|_{\rho-2\delta} \le 2m_6, \quad \|(N^{(1)})^{-1}\|_{\rho-7\delta} \le 8m_6,$$

where, to apply Neumann's series, we have used that (138) implies $\|N^{-1}\Delta\bar{N}\|_{\rho-2\delta} \le 1/2$ and $\|Db\|_{\rho-7\delta} \le 1/2$. To improve the bound above for $(N^{(1)})^{-1}$, we write:

$$
\begin{aligned}
N^{-1} - (N^{(1)})^{-1} &= (N^{(1)})^{-1}(N^{(1)} - N)N^{-1}, \\
N^{(1)} - N &= \bar{N}^{(1)}(\theta + b) - N + \bar{N}^{(1)}(\theta + b)Db + (Db)^\top \bar{N}^{(1)}(\theta + b)(\mathrm{Id}_r + Db), \\
\bar{N}^{(1)}(\theta + b) - N &= \Delta\bar{N}(\theta + b) + N(\theta + b) - N, \\
N_{i,j}(\theta + b(\theta)) - N_{i,j}(\theta) &= \left(\int_0^1 DN_{i,j}(\theta + sb(\theta))\,ds\right)b(\theta), \quad i,j = 1,\dots,r.
\end{aligned}
$$

Then, we establish the following bounds (we use that (138) implies that $m/(\gamma\delta^{\nu+1}) \le m_6 - \|N^{-1}\|_\rho$):

$$\|DN_{i,j}\|_{\rho-\delta} \le \frac{m}{\delta}, \quad \|N(\theta + b) - N\|_{\rho-7\delta} \le m\frac{\mu}{\gamma\delta^{\nu+1}}, \quad \|\bar{N}^{(1)}(\theta + b) - N\|_{\rho-7\delta} \le m\frac{\mu}{\gamma\delta^{\nu+1}},$$

$$\|N^{(1)} - N\|_{\rho-7\delta} \le m\frac{\mu}{\gamma\delta^{\nu+1}}, \quad \|N^{-1} - (N^{(1)})^{-1}\|_{\rho-7\delta} \le m\frac{\mu}{\gamma\delta^{\nu+1}}, \quad \|(N^{(1)})^{-1}\|_{\rho-7\delta} < m_6.$$

We define $S^{(1)}$ in terms of $\tau^{(1)}$ analogously as S is defined in terms of τ in (14):

$$S^{(1)} = (N_+^{(1)})^{-1}(D\tau_+^{(1)})^\top DF(\tau^{(1)})\,JD\tau^{(1)}\,(N^{(1)})^{-1}.$$

To verify theat $\det(\langle S^{(1)}\rangle_\theta) \ne 0$, we control the size of $S^{(1)} - S$ in terms of the corresponding bounds for N^{-1}, $(N^{(1)})^{-1}$, $D\tau$, $D\tau^{(1)}$, $DF(\tau)$, $DF(\tau^{(1)})$, $(N^{(1)})^{-1} - N^{-1}$, $D\tau^{(1)} - D\tau$, and $DF(\tau^{(1)}) - DF(\tau)$. This last difference is the only object in this list that we have not previously addressed. We write:

$$DF_j(\tau^{(1)}) - DF_j(\tau) = \left(\int_0^1 D^2F_j(\tau + s(\tau^{(1)} - \tau))\,ds\right)(\tau^{(1)} - \tau), \quad j = 1,\dots,2r,$$

and we conclude the following estimates:

$$\|DF(\tau^{(1)}) - DF(\tau)\|_{\rho-7\delta} \le m\frac{\mu}{\gamma\delta^\nu}, \quad \|S^{(1)} - S\|_{\rho-8\delta} \le m\frac{\mu}{\gamma\delta^{\nu+1}}, \quad |\langle S^{(1)}\rangle_\theta - \langle S\rangle_\theta| \le m\frac{\mu}{\gamma\delta^{\nu+1}}.$$

Next, we write $\langle S^{(1)}\rangle_\theta$ as follows:

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

We apply to expression above the estimates on Neumann's series displayed in (52). Explicitly, we observe that $|(\langle S\rangle_\theta)^{-1}(\langle S^{(1)}\rangle_\theta - \langle S\rangle_\theta)| \le m\mu/(\gamma\delta^{\nu+1}) < 1/2$ and that $m\mu/(\gamma\delta^{\nu+1}) < m_9 - |(\langle S\rangle_\theta)^{-1}|$ (see (138)), and we conclude that $\langle S^{(1)}\rangle_\theta$ is non-singular and that the following bounds are met:

$$|(\langle S^{(1)}\rangle_\theta)^{-1} - (\langle S\rangle_\theta)^{-1}| \le m\frac{\mu}{\gamma\delta^{\nu+1}}, \qquad |(\langle S^{(1)}\rangle_\theta)^{-1}| < m_9.$$

To finish this part of the proof we should control the new errors $\{e_j^{(1)}\}_{j=1}^3$. Regarding to $e_1^{(1)}$ and $\nabla e_1^{(1)}$, we should consider the following bounds (see equations (92), (93), (95), (97), (98), (99), (100), (102), (103), and (104)):

$$\|R_2\|_{\rho-\delta} \leq m\mu^2, \quad \|\bar{R}_2\|_{\rho-3\delta} \leq m\frac{\mu^2}{\delta}, \quad \|\hat{e}_1\|_{\rho-3\delta} \leq m\frac{\mu^2}{\delta}, \quad \|\hat{e}_2\|_{\rho-7\delta} \leq m\frac{\mu^2}{\delta}, \quad \|\hat{e}_3\|_{\rho-7\delta} \leq m\frac{\mu^2}{\delta},$$

$$\|\hat{e}_4\|_{\rho-7\delta} \leq m\frac{\mu^2}{\delta}, \quad \|\hat{\alpha}\|_{\rho-3\delta} \leq m\frac{\mu}{\delta}, \quad \|\hat{\beta}\|_{\rho-3\delta} \leq m\frac{\mu}{\delta}, \quad \|\hat{e}_5\|_{\rho-3\delta} \leq m\frac{\mu^2}{\delta}, \quad \|e_1^{(1)}\|_{\rho-7\delta} \leq m\frac{\mu^2}{\delta}.$$

To derive bounds above, we have used that $\|D^2\tau_j\|_{\rho-\delta} \leq m/\delta$, $\forall j = 1, \ldots, 2r$, that $\mu/\delta \leq 1$, as well as the inclusions of sets in (140) to control the involved compositions. Then, we conclude that $\|\nabla e_1^{(1)}\|_{\rho-8\delta} \leq m\mu^2/\delta^2$.

For $e_2^{(1)}$ we consider the following bounds (see (106), (111), (113), (117), (118), (120), (122), (124), and (125)):

$$\|c\|_{\rho-\delta} \leq m\mu, \quad \|Dc\|_{\rho-2\delta} \leq m\frac{\mu}{\delta}, \quad \|\hat{c}\|_{\rho} \leq m\mu, \quad \|D\hat{c}\|_{\rho-\delta} \leq m\frac{\mu}{\delta}, \quad \|\hat{e}_6\|_{\rho-2\delta} \leq m\frac{\mu^2}{\delta},$$

$$\|DR_1\|_{\rho-2\delta} \leq m\frac{\mu}{\delta}, \quad \|D\bar{R}_2\|_{\rho-4\delta} \leq m\frac{\mu^2}{\delta^2}, \quad \|\hat{e}_7\|_{\rho-4\delta} \leq m\frac{\mu^2}{\delta^2}, \quad \|\hat{R}_1\|_{\rho-7\delta} \leq m\frac{\mu}{\delta}, \quad \|\hat{R}_2\|_{\rho-7\delta} \leq m\frac{\mu^2}{\delta^2},$$

$$\|\hat{e}_9\|_{\rho-7\delta} \leq m\frac{\mu^2}{\delta^2}, \quad \|\hat{e}_{10}\|_{\rho-4\delta} \leq m\frac{\mu^2}{\delta^2}, \quad \|\hat{e}_8\|_{\rho-4\delta} \leq m\frac{\mu^2}{\delta^2}, \quad \|e_2^{(1)}\|_{\rho-7\delta} \leq m\frac{\mu^2}{\delta^2}.$$

To obtain some of the above bounds we have used again the inclusions of sets in (140). Moreover, we have also used that if we introduce the vectors $\bar{f} = (D\bar{\tau}^{(1)})^\top J \bar{\tau}^{(1)}$ and $f = (D\tau)^\top J \tau$, then we have, for any $j = 1, \ldots, 2r$:

$$\|\bar{f}_j\|_{\rho-2\delta} \leq m, \quad \|D\bar{f}_j\|_{\rho-3\delta} \leq \frac{m}{\delta}, \quad \|D^2\bar{f}_j\|_{\rho-4\delta} \leq \frac{m}{\delta^2}, \quad \|\bar{f} - f\|_{\rho-2\delta} \leq m\frac{\mu}{\delta}, \quad \|D(\bar{f} - f)\|_{\rho-3\delta} \leq m\frac{\mu}{\delta^2},$$

where we observe that $\bar{f}_j = \langle \partial_{\theta_j}\bar{\tau}^{(1)}, J\bar{\tau}^{(1)}\rangle$ and that $\bar{f} - f = (D\bar{\tau}^{(1)})^\top J \Delta\tau - (D(\Delta\tau))^\top J\tau$. Finally, to control $\hat{e}_8$ we use the estimate $\|\nabla(\tau_+^\top J \hat{e}_5)\|_{\rho-3\delta} \leq m\mu^2/\delta^2$.

Regarding to $e_3^{(1)}$, we consider the following bounds, where the one for $D^2\bar{\tau}_j^{(1)}$ holds $\forall j = 1, \ldots, 2r$ (see (128), (130), (131), (132), (133), (134), the explicit expression of the equation for a in (41), and the inclusions in (140)):

$$\|\mathcal{L}_\omega a\|_\rho \leq \mu_1 + \mu_2, \quad \|D^2\bar{\tau}_j^{(1)}\|_{\rho-3\delta} \leq \frac{m}{\delta}, \quad \|\hat{e}_{11}\|_{\rho-3\delta} \leq m\frac{\mu^2}{\delta}, \quad \|\hat{e}_{12}\|_{\rho-7\delta} \leq m\frac{\mu^2}{\delta},$$

$$\|\hat{e}_{13}\|_{\rho-7\delta} \leq m\frac{\mu}{\delta}, \quad \|\hat{e}_{14}\|_{\rho-3\delta} \leq m\frac{\mu^2}{\delta}, \quad \|\hat{e}_{15}\|_{\rho-7\delta} \leq m\frac{\mu^2}{\delta}, \quad \|e_3^{(1)}\|_{\rho-7\delta} \leq m\frac{\mu^2}{\delta}.$$

Remark 4.2. *This bound for $e_3^{(1)}$ is much better than was expected a priori. The much worse estimate $\|e_3^{(1)}\|_{\rho-7\delta} \leq \mu^2/(\gamma\delta^{\nu+2})$ would have lead to the same convergence condition (20) in theorem 2.6. This so good bound for $e_3^{(1)}$ is due to the fact that the contribution of b to $e_3^{(1)}$ only enters through compositions (since we are solving the nonlinear equation for b in (41)). The reader can compare with the proof of theorem 2.6 in [31], for the Hamiltonian case, in which the equation for b is a linear cohomological equation (so the final estimate for $e_3^{(1)}$ is much worse).*

We set the value of the constant m as one for which all the above estimates are met. We also introduce:

$$\rho^{(1)} = \rho - 8\delta, \qquad \mu_1^{(1)} = \mu_2^{(1)} = m\frac{\mu^2}{\delta^2}, \qquad \mu_3^{(1)} = m\frac{\mu^2}{\delta},$$

$$\Lambda = \max\{\|D\tau^{(1)} - D\tau\|_{\rho^{(1)}}, \|(N^{(1)})^{-1} - N^{-1}\|_{\rho^{(1)}}, |(\langle S^{(1)}\rangle_\theta)^{-1} - (\langle S\rangle_\theta)^{-1}|\}. \tag{141}$$

The summary of computations above is as follows. Suppose that τ is a parameterization of a quasi-torus of F for which the hypotheses of theorem 2.6 are fulfilled, and that μ verifies (138) for the selected δ. We compute a new parameterization $\tau^{(1)}$ by the application of the method introduced in definition 3.6. Then, we have shown that the

inequalities (12), (15), (19) of the statement of theorem 2.6 still hold if we replace τ, N, S, e, $\{e_j\}_{j=1}^3$, ρ, and $\{\mu_j\}_{j=1}^3$ by $\tau^{(1)}$, $N^{(1)}$, $S^{(1)}$, $e^{(1)}$, $\{e_j^{(1)}\}_{j=1}^3$, $\rho^{(1)}$, and $\{\mu_j^{(1)}\}_{j=1}^3$, respectively. Additionally, we also have:

$$\|\tau^{(1)} - \tau\|_{\rho^{(1)}} \le m\frac{\mu}{\gamma\delta^\nu}, \qquad \Lambda \le m\frac{\mu}{\gamma\delta^{\nu+1}}, \tag{142}$$

Iterative application of the method. To finish the proof, it only remains to show that the iterative application of the process to the initial parameterization τ converges to τ^*. To do that, we denote by $\{\tau^{(n)}\}_{n\ge 0}$ the sequence of parameterizations obtained by iteration of $\tau^{(0)} = \tau$. Similarly, we also denote by $\tau^{(n)}$, $N^{(n)}$, $S^{(n)}$, $e^{(n)}$, $\{e_j^{(n)}\}_{j=1}^3$, the corresponding objects associated to $\tau^{(n)}$. We also introduce the following recurrences, $\forall n \ge 1$:

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

$$\delta^{(n)} = \frac{\delta^{(n-1)}}{2}, \quad \rho^{(n)} = \rho^{(n-1)} - 8\delta^{(n-1)}, \quad \mu_1^{(n)} = \mu_2^{(n)} = m\frac{(\mu^{(n-1)})^2}{(\delta^{(n-1)})^2}, \quad \mu_3^{(n)} = m\frac{(\mu^{(n-1)})^2}{\delta^{(n-1)}},$$

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

where

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

We observe that $\lim_{n\to\infty}\rho^{(n)} = \rho - 16\delta = \rho^*$ and that $\lim_{n\to\infty}\mu^{(n)} = 0$, since condition (20) means that $\chi < 1$. Furthermore, we denote by $\Theta^{(n)}$ the quantity defined as Θ in (16), by replacing τ, N, S, and ρ, by $\tau^{(n)}$, $N^{(n)}$, $S^{(n)}$, and $\rho^{(n)}$, respectively. Similarly, $\Lambda^{(n)}$ is defined as Λ in (141), now by replacing τ, $\tau^{(1)}$, N, $N^{(1)}$, S, $S^{(1)}$, and $\rho^{(1)}$, by $\tau^{(n)}$, $\tau^{(n+1)}$, $N^{(n)}$, $N^{(n+1)}$, $S^{(n)}$, $S^{(n+1)}$, and $\rho^{(n+1)}$, respectively. In particular, $\Theta^{(0)} = \Theta$ and $\Lambda^{(0)} = \Lambda$.

We proceed by induction with respect to $n \ge 1$. Using notations above and the bounds for the first iteration, we conclude that we can iterate n times provided that the following bounds are met (see (137) and (138)):

$$m\frac{\mu^{(j)}}{\gamma(\delta^{(j)})^{\nu+1}} < \min\left\{\frac{1}{2}, \frac{m_8}{m_5 \cdot m_6}, \Theta^{(j)}\right\}, \qquad j = 0, \dots, n-1. \tag{143}$$

We will verify that (143) holds for any $n \ge 1$. Using that $\chi < 1$ and condition (20) we obtain (see defintion (137)):

$$m\frac{\mu^{(j)}}{\gamma(\delta^{(j)})^p} \le m\frac{\mu}{\gamma\delta^p}2^{j(p-\nu-2)}\chi^{2^j-1} \le m\frac{\mu}{\gamma\delta^p}\left(\frac{1}{2}\right)^j \le m\frac{\mu}{\gamma\delta^p} < \min\left\{\frac{1}{2}, \frac{m_8}{m_5 \cdot m_6}\right\}, \qquad \forall j \ge 0,$$

$$\sum_{j=0}^\infty m\frac{\mu^{(j)}}{\gamma(\delta^{(j)})^p} \le m\frac{\mu}{\gamma\delta^p}\sum_{j=0}^\infty\left(\frac{1}{2}\right)^j \le 2m\frac{\mu}{\gamma\delta^p} < \Theta, \tag{144}$$

where $p \in \{\nu, \nu+1\}$. In particular, we conclude that bounds in (143) involving $1/2$ and $m_8/(m_5 \cdot m_6)$ are fulfilled for any $j \ge 0$. Next, we start the induction process. The fist step is immediate, since condition (20) implies that (143) holds for $n = 1$ (i.e., for $j = 0$). To perform the general case, we assume that (143) is satisfied up to some $n \ge 1$. This fact means that we can iterate the process n times and that we have (see (142)):

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

In particular, using the definitions of $\Lambda^{(j)}$ and bounds above, we obtain:

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

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

Consequently, the definition of Θ in (16) implies:

$$m\frac{\mu^{(n)}}{\gamma(\delta^{(n)})^{\nu+1}} < \|D\tau\|_\rho - \|D\tau^{(n)}\|_{\rho^{(n)}} + \Theta \leq m_5 - \|D\tau^{(n)}\|_{\rho^{(n)}}.$$

Performing analogous computations as above for $\|\tau^{(n)}\|_{\rho^{(n)}}$, $\|(N^{(n)})^{-1}\|_{\rho^{(n)}}$, $d_{\tau^{(n)},\rho^{(n)},\mathcal{U}}$, and $|(\langle S^{(n)}\rangle_\theta)^{-1}|$, it is immediate to verify that the bounds in (143) involving $\Theta^{(j)}$ are also met for $j = n$. Note that the cases $\|\tau^{(n)}\|_{\rho^{(n)}}$ and $d_{\tau^{(n)},\rho^{(n)},\mathcal{U}}$ are slightly different (but also straightforward) since then we use (144) for $p = \nu$.

Hence, we have verified that we can iteratively apply to τ the quasi-Newton-like method of definition 3.6 infinitely many times. Since $\lim_{n\to\infty}\mu^{(n)} = 0$, the results of the statement of theorem 2.6 follow by taking the limit as $n \to \infty$ of this process. Specifically, there is $\tau^* = \lim_{n\to\infty}\tau^{(n)}$, which is analytic in $\Delta(\rho^*)$, and solves the invariance equation $F(\tau^*) = \tau_+^*$. This last assertion is immediate since $\lim_{n\to\infty}\mu^{(n)} = 0$ means not only that $e_j^* = 0$, $j = 1, 2, 3$, but also that $e^* = 0$ (so, at this point we do not need to use proposition 3.2). Furthermore:

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

and, similarly, $\|D\tau^* - D\tau\|_{\rho^*} \leq 2m\mu/(\gamma\delta^{\nu+1})$.

References

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

[2] H.W. Broer and M.B. Sevryuk. KAM theory: Quasi-periodicity in dynamical systems. In *Handbook of Dynamical Systems*, volume 3C, pages 249–344. Elsevier, 2010.

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

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

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

[6] R. de la Llave. 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, RI, 2001.

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

[8] J.-Ll. Figueras, À. Haro, and A. Luque. Rigorous computer-assisted application of KAM theory: A modern approach. *Foundations of Computational Mathematics*, 17(5):1123–1193, 2017.

[9] E. Fontich, R. de la Llave, and Y. Sire. Construction of invariant whiskered tori by a parameterization method. Part I: Maps and flows in finite dimensions. *J. Differential Equations*, 246:3136–3213, 2009.

[10] A. González and R. de la Llave. Analytic smoothing of geometric maps with applications to KAM theory. *J. Differential Equations*, 245(5):1243–1298, 2008.

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

[12] À. Haro. An algorithm to generate canonical transformations: application to normal forms. *Phys. D*, 167(3-4):197–217, 2002.

[13] À. Haro, M. Canadell, J.-Ll. Figueras, A. Luque, and J.M. Mondelo. *The parameterization method for invariant manifolds*, volume 195 of *Applied Mathematical Sciences*. Springer, [Cham], 2016. From rigorous results to effective computations.

[14] À. Haro and R. de la Llave. A parameterization method for the computation of invariant tori and their whiskers in quasi-periodic maps: numerical algorithms. *Discrete Contin. Dyn. Syst. Ser. B*, 6(6):1261–1300, 2006.

[15] À. Haro and R. de la Llave. A parameterization method for the computation of invariant tori and their whiskers in quasi-periodic maps: rigorous results. *J. Differential Equations*, 228(2):530–579, 2006.

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

[17] À. Jorba, R. de la Llave, and M. Zou. Lindstedt series for lower-dimensional tori. In *Hamiltonian systems with three or more degrees of freedom (S'Agaró, 1995)*, volume 533 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 151–167. Kluwer Acad. Publ., Dordrecht, 1999.

[18] A.N. Kolmogorov. 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 p. 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.

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

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

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

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

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

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

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

[26] H. Rüssmann. 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)*, pages 598–624. Lecture Notes in Phys., Vol. 38. Springer, Berlin, 1975.

[27] H. Rüssmann. 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.

[28] H. Rüssmann. On optimal estimates for the solutions of linear difference equations on the circle. In *Proceedings of the Fifth Conference on Mathematical Methods in Celestial Mechanics (Oberwolfach, 1975), Part I. Celestial Mech.*, volume 14, 1976.

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

[30] J. Villanueva. Kolmogorov theorem revisited. *J. Differential Equations*, 244(9):2251–2276, 2008.

[31] J. Villanueva. A new approach to the parameterization method for lagrangian tori of hamiltonian systems. *J. Nonlinear Sci.*, 27(2):495–530, 2017.

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