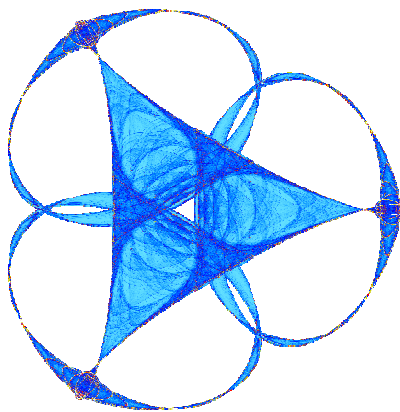# COMPUTATION OF LIMIT CYCLES AND THEIR ISOCHRONES: FAST ALGORITHMS AND THEIR CONVERGENCE

By

**Gemma Huguet and Rafael de la Llave**

**IMA Preprint Series #2415**

(December 2012)

# COMPUTATION OF LIMIT CYCLES
# AND THEIR ISOCHRONES: FAST ALGORITHMS
# AND THEIR CONVERGENCE

GEMMA HUGUET AND RAFAEL DE LA LLAVE

ABSTRACT. In this paper we develop efficient algorithms to compute limit cycles and their isochrones (i.e. the sets of points with the same asymptotic phase) for planar vector fields. We formulate a functional equation for the parameterization of the invariant cycle and its isochrones and we show that it can be solved by means of a Newton method. Using the right transformations, we can solve the equation of the Newton step efficiently.

The algorithms are efficient in the sense that if we discretize in $N$ points, a Newton step requires $O(N)$ storage and $O(N \log(N))$ operations (in Fourier discretization) or $O(N)$ operations in other discretizations.

We prove convergence of the algorithms and present a validation theorem in an *a-posteriori* format. That is, we show that if there is an approximate solution of the invariance equation that satisfies some some mild non-degeneracy conditions, then, there is a true solution nearby. Thus, our main theorem can be used to validate numerically computed solutions.

The theorem also shows that the isochrones are analytic and depend analytically on the base point. Moreover, it establishes smooth dependence of the solutions on parameters and provides efficient algorithms to compute perturbative expansions with respect to external parameters.

We include a discussion on the numerical implementation of the algorithms as well as numerical results for a representative example.

## 1. INTRODUCTION

In recent times, there has been increased interest into the asymptotic phase near a limit cycle because of its applications to Neuroscience. The sets with constant asymptotic phase, called isochrones [Win75]

---

and the change of phase under perturbation called the "phase resetting curve" (PRC) play an important role in the understanding of the response of neurons to brief stimuli and its synchronization properties [Izh07, Erm96, MC08]. The relation of isochrones and PRCs to the theory of normally hyperbolic manifolds was pointed out in [Guc75]. Since then, there has been a large effort into computing isochrones and PRCs. Recent papers close to our goal are [GH09] which describes and implements computations of isochrones and PRCs even for points which are not on the limit cycle, [TF10] which presents algorithms to all orders, [SG09] which describes and implements algorithms for bursting neurons and [OM10] which describes and implements algorithms for the computation of isochrones suitable for very stiff systems.

In this paper we describe efficient algorithms for the computation of isochrones and PRCs, that allow us to obtain infomation about the phase not just in an infinitesimal neighborhood of the limit cycle but also in a whole neighborhood. The core mathematical result of this paper is Theorem 3.2, which provides estimates for the convergence of the algorithm. In addition to this, Theorem 3.2 shows that the isochrones are analytic and depend analytically on the base point, establishes smooth dependence of the solutions on parameters and provides with efficient algorithms to compute perturbative expansions with respect to external parameters.

The starting point of the mathematical formulation of the problem is the work [Guc75] which pointed out that isochrones are just the stable manifolds of a point in the limit cycle in the sense of the theory of normally hyperbolic manifolds. Hence, following [CFL03a, CFL03b, CFL05], we formulate the problem as solving a functional equation (see (4)) for the parameterization of the invariant circle and its isochrones. This method was also used in [GH09], where the invariance equation was solved using a step by step method. The main novelty here is that we apply a Newton method to solve this functional equation. The key to obtain both an efficient algorithm and good estimates that lead to convergence, is to use several identities and algebraic manipulations to transform the equation for the Newton step into an equation which can be solved efficiently.

These identities, which have an interesting geometric interpretation, are obtained by taking derivatives of the invariance equation. As it turns out, the Newton step involves some *"loss of derivatives"*, i.e. the norm of the remainder after a Newton step is controled by the square of the norm of derivatives of the remainder before the Newton step. It is well known since [Kol54, Mos66b, Mos66a] that the quadratic estimates in one step lead to convergence even when there is loss of derivatives.

Of course, numerical implementations exhibit quadratic convergence irrespective of the fact that to establish it one uses a sophisticated method.

The algorithms we present can be implemented rather straightforwardly using a package manipulating Fourier-Taylor series of the type that is commonly used in celestial mechanics [BG69, Har08, LG12]. The algorithm is highly efficient in the sense that if we discretize the function using $N$ terms of the Fourier-Taylor series, then a Newton step requires $O(N)$ storage and $O(N \log N)$ operations but it has the quadratic convergence of the Newton algorithm (i.e. after an application of the algorithm, the error is roughly the square of the original error). If we discretize in splines, or using collocation methods, using $N$ points, a Newton step requires $O(N)$ storage and $O(N)$ operations (with a larger constant).

We have implemented these algorithms using Fourier series and the results in some representative cases are reported in Section 8. The algorithm is highly accurate (in the example considered, one can get the isochrones up to a thousand times of the roundoff error) and it also gives information on the derivatives of the isochrones. The computations in Section 8 run in seconds in a standard laptop.

We call attention that the main Theorem 3.2 is formulated in the *"a-posteriori"* format. That is, Theorem 3.2 shows that given an approximate solution of the invariance equation (4), which satisfies some explicit non-degeneracy conditions, then, there is a true solution nearby. Theorems in *"a-posteriori"* format can be used to validate the computations and allow us to be confident of the calculations even when they are close to breakdown. Hence, the calculations obtained using Algorithm 4.3 can easily be turned into computer assisted proofs. It suffices to estimate rigorously the error and the condition numbers.

From the analytic point of view, we stress that the method leads also to several other consequences, such as uniqueness and smooth dependence on parameters. Some of them can be obtained by the standard methods of ODE (notably the smooth dependence on parameters for the limit cycle). However, the regularity of isochrones does not seem to follow from the general theory of normal hyperbolicity. We also hope that the methods developed here can be extended to other problems.

This paper is organized as follows: in Section 2, we formulate the problem as a functional equation. In Section 3, we state the main result (Theorem 3.2). In Section 4 we describe the iterative step that will be used to improve approximations for $K$. In particular, in Section 4.4, we present the iterative step in an algorithmic form (Algorithm 4.3). This iterative step will be used in Section 5 to prove that the method

converges in some appropriate norms defined in 3.1. In Section 6 we present some consequences of the formalism such as smooth dependence on parameters. The Algorithm 4.3 will also be discussed in Section 7 where we supplement it with implementation details. In Section 8 we present numerical results for a representative example.

Note that the tools and standards of the sections of this paper are very different and could appeal to different communities. Hence, we have striven to make the specific sections readable independently. For instance, in Sections 4.4, 7 and 8, we use the language of algorithm theory, and discuss storage, operation counts, etc. In contrast, in Sections 5 and 6 we use rigorous mathematical estimates.

## 2. Set up of the problem

We consider a differential equation in the plane

$$(1) \qquad\qquad \dot{x} = X(x) \qquad x \in \mathbb{R}^2,$$

and denote by $X^t$ the flow associated to (1). That is, $X^t(x_0)$ solves $\frac{d}{dt}X^t(x_0) = X(X^t(x_0))$, $X^0(x_0) = x_0$. We assume that (1) admits a hyperbolic limit cycle and that $X$ is analytic.

More precisely, we assume for some map $K_0 : \mathbb{T} \to \mathbb{R}^2$, that $x(t) = K_0(\omega t)$ is a solution of (1) and, furthermore that $\mathcal{K} = K_0(\mathbb{T}^1)$ is an exponentially attracting set. That is, if $y$ is close enough to $\mathcal{K}$,

$$(2) \qquad\qquad d(X^t(y), \mathcal{K}) \le C\, e^{-\lambda t},$$

for some $C, \lambda > 0$.

Given (2), as pointed out in [Guc75], it follows that given such a $y$, we can find a unique $\Phi(y)$ in such a way that

$$(3) \qquad\qquad |X^t(y) - K_0(t + \Phi(y))| \le C\, e^{-\lambda t}.$$

The goal is to show (see Theorem 3.2 for a more precise statement) that, in these circumstances, we can find $K : \mathbb{T}^1 \times [-1, 1] \to \mathbb{R}^2$, an analytic local diffeomorphism, such that

$$(4) \qquad\qquad X \circ K(\theta, s) = DK(\theta, s) \begin{bmatrix} \omega \\ \lambda s \end{bmatrix}.$$

Using the more concise notation $A_{\omega,\lambda} = \begin{bmatrix} \omega \\ \lambda s \end{bmatrix}$, equation (4) can be written as

$$X \circ K(\theta, s) = DK(\theta, s) A_{\omega,\lambda}.$$

Equation (4) will be the centerpiece of our approach. Note that, if (4) holds, (2) also holds with the same value of $\lambda$. Of course, (2) holds for

a range of $\lambda$'s while (4) determines $\lambda$ in a unique way. From now on, we will use $\lambda$ to denote the optimal $\lambda$ in (2). [1]

We should think of (4) as a functional equation for the mapping $K$ and for the numbers $\omega, \lambda$. The vector field $X$ is, of course, known.

Note that we are taking the convention that the functions are 1-periodic, not $2\pi$-periodic. Hence, $\omega$ is the inverse of the period.

Note that with our conventions $\lambda$ is positive when the limit cycle is repulsive and negative when the limit cycle is stable. The method works in both cases. In the stable case, the isochrones are obtained by fixing the asymptotic phase in the future. In the unstable case, the isochrones are obtained by fixing the asymptotic phase in the past. Of course, one can pass from the stable case to the unstable case just by changing the direction of time (equivalently, the sign of the vector field $X$).

2.1. **Geometric interpretation of invariance equation** (4). We can think of (4) as a change of variables that turns the vector field $X$ into the straight vector field on $\mathbb{T}^1 \times [-1, 1]$

$$(5) \qquad A_{\omega\lambda} \equiv \begin{bmatrix} \omega \\ \lambda s \end{bmatrix}.$$

We can also write (4) as

$$X \circ K(\theta, s) = [\omega\partial_\theta + \lambda s\partial_s] K(\theta, s),$$

but the formulation as in (4) is geometrically more natural.

It is straightforward to show that if we have (4), the evolution in the coordinates $(\theta, s)$ becomes

$$(6) \qquad X^t(K(\theta, s)) = K(\theta + t\omega, s\,e^{\lambda t}).$$

Indeed, note that, using (4) we have:

$$\frac{d}{dt}K(\theta + t\omega, s\,e^{\lambda t}) = (\omega\partial_\theta + s\,\lambda\,e^{\lambda t}\partial_s)K(\theta + t\omega, s\,e^{\lambda t})$$
$$= X\left(K(\theta + t\omega, s\,e^{\lambda t})\right).$$

We can describe (6) as saying that if we perform the change of variables given by $K$, the coordinates $(\theta, s)$ evolve by the linearized evolution

$$(7) \qquad \Lambda^t(\theta, s) = (\theta + \omega t, s\,e^{\lambda t}),$$

that is, we have $X^t \circ K = K \circ \Lambda^t$.

---

[1]one of the consequences of the theory developed here is that the optimal $\lambda$ in (2) exists. That is, there are no subexponential corrections. See Section 2.3 for more details.

In particular, the isochrones are just the sets obtained by fixing $\theta$ and letting $s$ vary

$$(8) \qquad \mathcal{S}_\theta = \{K(\theta, s) \mid s \in [-1, 1]\} \ .$$

In Theorem 3.2, we will show that $K$ is analytic, and, as a corollary, that $\mathcal{S}_\theta$ are analytic manifolds and that they depend analytically on $\theta$.

Note that if we have (4),

$$K_0(\theta) = K(\theta, 0)$$

is a limit cycle, so that we see that the isochrones $\mathcal{S}_\theta$ are curves transversal to the limit cycle. Indeed $K_1(\theta)$ is the tangent to the isochrone at $K_0(\theta)$. In symbols,

$$K_1(\theta) = \partial_s K(\theta, s)_{|s=0}.$$

Note that even if the foliation by isochrones is invariant, the individual leaves are not invariant. Indeed, we have

$$(9) \qquad \mathcal{S}_{\theta+\omega t} = X^t(\mathcal{S}_\theta).$$

*Remark* 2.1. It has been known since Poincaré [Poi78] that one can linearize in a neighborhood of a limit cycle for planar vector fields. If the vector field is analytic, the linearization is analytic. The linearization is just the function $K$ solving (4). Usually, the analyticity of $K$ is proved by reducing to a (one-dimensional) Poincaré map and then showing that the one-dimensional map can be conjugated to a linear one. The mathematical point of this paper is that we can study (4) directly and that approximate solutions for it can be validated. The method of study of the equation yields very efficient algorithms.

2.2. **Lack of uniqueness.** It is important to note that the solutions of (4) are never unique.

Indeed, for any $\theta_0 \in \mathbb{T}^1$ and $b \in \mathbb{R}$, if $(K, \omega, \lambda)$ is a solution of (4) and $\tilde{K}$ is defined by $\tilde{K}(\theta, s) = K(\theta + \theta_0, \ s \cdot b)$; then $(\tilde{K}, \omega, \lambda)$ is also a solution of (4).

We will show in Theorem 6.1 that this is the only source of nonuniqueness. In particular, $\omega$ and $\lambda$ are uniquely determined.

A practical consequence of this lack of uniqueness is that we can assume (by choosing $b$) that the domain of the parameter $s$ is $[-1, 1]$. This choice of parameters $b$ is convenient for theoretical calculations, but not essential. On the other hand, it is very important for numerical calculations. It is well known that the round-off becomes very problematic if we are working with numbers that are some orders of magnitude apart. By choosing $b$ appropriately, it is possible to aim for having the coefficients of the expansion of $K$ more or less constant in size so that roundoff is greatly reduced.

2.3. **Topological characterization of isochrones.** One of the consequences of (4) is that the isochrones admit a topological characterization. This is somewhat different from the characterization given by the theory or normally hyperbolic manifolds, which involves not only convergence, but convergence at a certain exponential rate. For simplicity, we present only the formula for the stable case, when $\lambda < 0$ and the asymptotic phase is the phase in the future.

Using (4) and the fact that $K$ is uniformly differentiable (so that it preserves rates of convergence up to a constant), we have for $0 < \eta \ll 1$:

(10)
$$
\begin{aligned}
W^s_{K(\theta,0)} &= \{P \in \mathbb{R}^2 \mid \ |X_t(P) - X_t(K(\theta,0))| \le C_{\eta,P} e^{-(|\lambda|-\eta)t} \quad \text{for all} \quad t \ge 0\} \\
&= \{K(\theta,s), s \in \mathbb{R}\} \\
&= \{P \in \mathbb{R}^2 \mid \ |X_t(P) - X_t(K(\theta,0))| \to 0 \quad \text{as} \quad t \to \infty\} \\
&= \{P \in \mathbb{R}^2 \mid \ |X_t(P) - X_t(K(\theta,0))| \le C_P e^{-|\lambda|t} \quad \text{for all} \quad t \ge 0\}
\end{aligned}
$$

The first characterization of $W^s_{K(\theta)}$ in (10) is the standard definition in normally hyperbolic theory. Note that it involves rates of convergence.

The third line of (10) is a purely topological characterization (it just involves convergence irrespective of the rate). The link between all the characterizations in (10) is using the characterization given in the second line of (10). The difference from the first and fourth lines of (10) is that in the first line we just assume that there is an open interval of rates of convegence. In the fourth line, we obtain that the end point of the rates of convergence allowed in the first line is also allowed.

The proof of the characterization follows from the observation that it is valid when the dynamics is $\Lambda^t$ and that changing variables by $K$ does not change the rates of convergence.

The characterization above shows that, for the problem at hand, orbits that converge, converge exponentially fast and that the exponential rate is always the optimal one (i.e. there are no polynomial corrections). This result depends very much on the low dimensionality of the problem and is false in more general normally hyperbolic manifolds.

## 3. STATEMENT OF THE MAIN ANALYTICAL RESULT, THEOREM 3.2

To formulate precisely Theorem 3.2, we need some definitions of norms in which to measure functions.

3.1. **Definition of norms.** We use the standard supremum norms in KAM theory. They seem to give the sharpest results in loss of differentiabiliy. As for the numerical work with them, see Section 7.4.

We denote

$$\mathbb{T}_\rho = \{\theta \in \mathbb{C}/\mathbb{Z} \,|\, |\mathrm{Im}(\theta)| \le \rho\}$$

(11)
$$B_\beta = \{s \in \mathbb{C} \,|\, |s| \le \beta\}$$

$$\mathcal{D}_{\beta,\rho} = \mathbb{T}_\rho \times B_\beta$$

**Definition 3.1.** *Given a periodic function* $f(\theta) = \sum_{k\in\mathbb{Z}} \hat{f}_k e^{2\pi ik\theta}$ *and a number* $\rho > 0$, *we define*

(12)
$$\|f\|_\rho = \sup_{\theta\in\mathbb{T}_\rho} |f(\theta)|.$$

*Given a family of periodic functions* $K(\theta, s) = \sum_{n\in\mathbb{N}} K_n(\theta) s^n$ *with* $K_n(\theta)$ *periodic and two numbers* $\rho, \beta > 0$ *we define*

(13)
$$\|K\|_{\beta,\rho} = \sup_{|s|\le\beta, \theta\in\mathbb{T}_\rho} |K(\theta, s)|.$$

Note that the definitions above are valid both in the cases that the functions are real valued or vector valued.

We consider the spaces $\mathcal{A}_\rho, \mathcal{A}_{\beta,\rho}$ consisting of the functions for which the above norms are finite. We consider them equipped with the corresponding norms, which makes $\mathcal{A}_\rho, \mathcal{A}_{\beta,\rho}$ Banach spaces.

Some elementary properties of these norms are presented in Section 5.1.

3.2. **Statement of Theorem 3.2.**

**Theorem 3.2.** *Let* $X$ *be an analytic vector field in a domain of* $\mathbb{R}^2$ *which extends to a domain* $\mathcal{C} \subset \mathbb{C}^2$.

*Assume that we can find an analytic parameterization* $K : \mathbb{T} \times [-1, 1] \to \mathbb{R}^2$ *and numbers* $\omega, \lambda$ *in such a way that:*

- $K(\mathbb{T}_\rho \times B_\beta) \subset \mathcal{C}$, $\mathrm{dist}(K(\mathbb{T}_\rho \times B_\beta), \mathbb{C}^2 - \mathcal{C}) = \zeta > 0$.
- $\|X \circ K - DKA_{\omega,\lambda}\|_{\beta,\rho} < \varepsilon$
- *For some* $0 < \delta < \rho/2$, *we have*

(14)
$$\varepsilon\delta^{-1}C \le 1$$

  *where* $C$ *is an explict expression developed in the proof depending on* $\sup_{x\in\mathcal{C}} |D^2 X|$, $\sup_{x\in\mathbb{T}_\rho\times B_\beta} |DK(x)|, |DK^{-1}(x)|, |D^2 K|$.

*Then, there exists* $K^*$ *an analytic local diffeomorphism and* $\omega^*, \lambda^* \in \mathbb{R}$ *such that*

(15)
$$X \circ K^*(\theta, s) = DK^*(\theta, s)A_{\omega^*,\lambda^*}.$$

*Furthermore,*

$$\|K - K^*\|_{\beta-\delta,\rho-\delta}, |\omega - \omega^*|, |\lambda - \lambda^*| \leq C\varepsilon. \tag{16}$$

Here, following the standard practice in KAM arguments, we denote by the letter $C$ quantities that depend on the quantities indicated after (16) even if the meaning could be different from line to line. In particular, the constant $C$ that appears in the conclusions (16) is different from the constant $C$ that appears in the hypothesis (14). There are several pairs of constants that work.

As we will see in subsequent sections, the proof of Theorem 3.2 is based in a rapidly convergent iteration. This iteration, which takes advantage of some geometric calculations, is not only the method of proof but also yields efficient algorithms.

The proof of Theorem 3.2 also leads to the conclusion that $K^*, \omega^*$ and $\lambda^*$ are locally unique (up to obvious choices of origins of coordinates in $K$ and scaling factors discussed in Section 2.2). We will also show that the solutions depend smoothly on parameters and that there are versions of the theorem for finitely differentiable vector fields. We discuss these results in more detail in Section 6.

In practice, the computation of $K$ using the algorithms justified by Theorem 3.2, provides extremely accurate representations of the isochrones in a neighborhood of the limit cycle. Once this accurate representation is known in a neighborhood, we can extend the computation using

$$\mathcal{S}_\theta = X^{-t}(\mathcal{S}_{\theta+\omega t}).$$

Note that if $\mathcal{S}_{\theta+\omega t}$ is known as a small curve, for $t > 0$, the isochrone $\mathcal{S}_\theta$ is much longer.

## 4. The iterative step for the computation of $(K, \omega, \lambda)$

In this section we will describe the computation of $(K, \omega, \lambda)$. More precisely, we will describe the procedure to obtain a more approximate solution out of a sufficiently approximate one. That is, we will describe the algebraic manipulations to simplify the invariance equation and we will state explicitly the functional equations that need to be solved as well as their solutions. We will present two methods to solve the equations needed (one using Fourier series and another one using integral representations).

This procedure will be the basis both of the convergence proof developed in Section 5 and of the description of the algorithmic steps in Section 4.4.

To obtain convergence proofs we need to supplement the discussion in this section with considerations of function space and estimates that quantify the assertion that the result of the step satisfies the invariance equation more accurately. For the numerical implementation, we need to supplement the discussion in this section with specifications on how to discretize functions and implement the elementary operations (algebraic operations, composition, derivatives, integrals, etc.) as well as how to solve the functional equations that appear in the Newton and quasi-Newton methods.

Of course, we will also need to show that the procedure to produce more accurate solutions can be iterated, hence obtaining a sequence of functions which converge to a solution.

## 4.1. **The Newton method.**

A convenient place to start the motivation of the iterative step to obtain a solution of (4) from approximate solutions is the Newton method.

Given an approximate solution $(K, \omega, \lambda)$ of (4)

$$(17) \qquad\qquad X \circ K - DK A_{\omega,\lambda} = E,$$

Newton method seeks an improved solution $K + \Delta$, $\omega + \sigma$, $\lambda + \eta$, in such a way that $\Delta, \sigma, \eta$ eliminates $E$ "in the linear approximation".

Since $X \circ (K + \Delta) \approx X \circ K + DX \circ K\Delta$ and $D(K + \Delta) = DK + D\Delta$, we have that the equation for the Newton method is

$$(18) \qquad DX \circ K\Delta - (D\Delta) A_{\omega,\lambda} - DK A_{\sigma,\eta} = -E.$$

One should think of (18) as an equation for $\Delta$ and $\eta, \sigma$ when all the other quantities are known. $X$ is given by the problem and $(K, \omega, \lambda)$ is the known approximation we are trying to improve.

Note that at this stage it is not clear that (18) has a solution because we have periodicity requirements on $K$. The fact that (18) has solutions will be established in Section 4.1. In Section 4.1 we will also provide quantitative estimates.

If we discretize our functions in some appropriate basis of funtions satisfying the periodicity conditions, equation (18) is a linear equation that can be solved using a linear solver. This is a reasonably practical approach in many circumstances and was used in [GH09].

In this paper, however, we will take a different approach. We will use several identities to obtain a change of variables which reduces (18) to a much simpler equation up to a certain error which is smaller than the original error and does not change the quadratic character of the Newton method. The use of these identities in linearization problems was pointed out in [Mos66b]. A more systematic study based on "group structure" of the equations is in [Zeh75]. Some extensions of

this approach were used in [LGJV05] for Hamiltonian systems, taking advantage of the geometric properties of the system. In our case, the geometric properties we take advantage of are, mainly, the lower dimensionality of the system.

4.2. **The quasi-Newton method.** We are given $(K, \omega, \lambda)$ satisfying (17). Notice that taking derivatives of (17) we obtain that our approximate solution $(K, \omega, \lambda)$ also satisfies

$$(19) \qquad (DX \circ K)DK - D^2KA_{\omega,\lambda} - DKDA_{\omega,\lambda} = DE,$$

where

$$DA_{\omega,\lambda} = \begin{pmatrix} 0 & 0 \\ 0 & \lambda \end{pmatrix}.$$

We emphasize that both (17) and, hence, (19) are information that is on hand at the beginning of the iterative step.

As we will see, one can use (19) to simplify the equation for the Newton method (18). The main idea is that rather than looking for $(\Delta, \sigma, \eta)$ in (18), we look for $(W, \sigma, \eta)$, where

$$(20) \qquad \qquad \Delta = DKW.$$

Note that if $DK$ is invertible, both $\Delta$ and $W$ are equivalent unknowns in the sense that if we know one, we can find the other one.

If we substitute (20) into (18) we obtain that (18) is equivalent to

$$(21) \quad DX \circ KDKW - D^2KWA_{\omega,\lambda} - DKDWA_{\omega,\lambda} - DKA_{\sigma,\eta} = -E.$$

Using (19), we obtain that (21) is equivalent to (recall that $D^2KWA_{\omega,\lambda} = D^2KA_{\omega,\lambda}W$ because $D^2K$ is a symmetric quadratic form):

$$(22) \qquad DKDA_{\omega,\lambda}W + (DE)W - DKDWA_{\omega,\lambda} - DKA_{\sigma,\eta} = -E.$$

The quasi-Newton method consists just in dropping the term $DEW$ from (22), which we argue, heuristically at the moment, is *"quadratically small"* because it is the product of two terms which are small (think of $W$ as of the same order of smallness than $E$). Of course, this heuristic idea that $DEW$ is small will be made rigorous when we perform estimates in Section 5.2.

Hence, we will consider the equation for $(W, \sigma, \eta)$,

$$(23) \qquad DKDA_{\omega,\lambda}W - DKDWA_{\omega,\lambda} - DKA_{\sigma,\eta} = -E$$

and then, consider the improved solution $(K + DKW, \omega + \sigma, \lambda + \eta)$. This will be referred to as the quasi-Newton step.

In the rest of this section we will just show that (23) has solutions. Estimates on the size of the solution and on the improvement of the error will be carried out in Section 5.2.

If we premultiply (23) by $DK^{-1}$ we obtain

(24) $$\begin{pmatrix} 0 & 0 \\ 0 & \lambda \end{pmatrix} W - DW A_{\omega,\lambda} - A_{\sigma,\eta} = -DK^{-1}E.$$

If we express (24) in components, using the shorthand $\tilde{E} = DK^{-1}E$, and denoting the components of $\tilde{E}$ and $W$ by subindices, $\tilde{E} = (\tilde{E}_1, \tilde{E}_2)$ and $W = (W_1, W_2)$, we obtain:

(25)
$$-(\omega\partial_\theta + \lambda s\partial_s)W_1 - \sigma = -\tilde{E}_1$$
$$\lambda W_2 - (\omega\partial_\theta + \lambda s\partial_s)W_2 - \eta s = -\tilde{E}_2.$$

The remarkable feature of (25) is that it only involves the linear operator with constant coefficients. As we will see next, these equations can be solved very efficiently either in Fourier coefficients ( see Lemma 4.1) or using explicit (and fast converging) integral formulas (see Lemma 4.2).

4.3. **Solution of the constant coefficients linearized equations.** In the this section, we will study the solvability of equations (25) both using (formal) Fourier series and improper (but rapidly convergent) integrals. Detailed estimates will be established in Section 5. Nevertheless, for the purpose of implementing algorithms, only the existence and the form of the solutions is needed.

4.3.1. *Fourier methods for the solutions of the quasi-Newton method.*

**Lemma 4.1.** *Consider given a formal series $\tilde{E} = \sum \tilde{E}_{j,k} s^j e^{2\pi ik\theta}$. Then, if $\tilde{E}_{00} = 0$, the equation for $u$*

(26) $$(\omega\partial_\theta + \lambda s\partial_s)u = \tilde{E}$$

*has the one dimensional family of formal series solutions*

$$\sum_{j\in\mathbb{N},\, k\in\mathbb{Z}} u_{j,k} s^j e^{2\pi ik\theta}$$

*with*

(27)
$$u_{j,k} = \frac{\tilde{E}_{j,k}}{2\pi i\omega k + \lambda j} \quad \text{if } (j,k) \neq (0,0)$$
$$u_{0,0} = \alpha$$

*for any $\alpha \in \mathbb{R}$. The solutions given in (27) are the only formal series solutions of (26). Furthermore if $\tilde{E}_{00} \neq 0$ there are no formal series solutions of (26).*

*If $\tilde{E}_{10} = 0$, then the equation for $u$*

(28) $$-\lambda u + (\omega\partial_\theta + \lambda s\partial_s)u = \tilde{E}$$

*has the one parameter family of formal series solutions*

$$\sum_{j\in\mathbb{N},\, k\in\mathbb{Z}} u_{jk} s^j e^{2\pi i k\theta}$$

*with*

(29)
$$u_{jk} = \frac{\tilde{E}_{jk}}{2\pi i\omega k + \lambda(j-1)} \quad \text{if } (j,k) \neq (1,0)$$

$$u_{1,0} = \alpha$$

*for any $\alpha \in \mathbb{R}$. The solutions given by (29) are the only formal series solutions of (28). Furthermore, if $\tilde{E}_{1,0} \neq 0$, (28) has no solutions.*

*Proof.* Taking Fourier series on both sides of (26) we have

$$u_{jk}(\lambda j + 2\pi i\omega k) = \tilde{E}_{jk}.$$

It is easy to see that $\lambda j + 2\pi i\omega k = 0$ if and only if $j = 0$, $k = 0$. Therefore, the solution $u$ is obtained by setting $u_{jk} = \tilde{E}_{jk}/(\lambda j + 2\pi i\omega k)$ when $(j,k) \neq (0,0)$ and $u_{00}$ is arbitrary.

Similarly, we observe that (28) is equivalent to

$$(\lambda j - \lambda + 2\pi i\omega k)u_{jk} = \tilde{E}_{jk}.$$

Again, we note that $\lambda(j-1) + 2\pi i\omega k = 0$ if and only if $j = 1$, $k = 0$ and then, the same argument as before applies. $\qquad\square$

Consider the expressions given by (27) and (29). If $\tilde{E}$ is not just a formal power series, but rather a smooth function, the solutions above will also have several regularity properties. The reason is that regularity of the error implies fast decay properties for the Fourier-Taylor coefficients, which in turn imply fast decay of the Fourier-Taylor coefficients of the solutions and, hence regularity properties of the solutions. Detailed estimates will be presented in Section 5.

The solutions of the equations (26) and (28) can be computed very efficiently when the functions are discretized in Fourier-Taylor series. Notice that if we store $N$ Fourier coefficients, we need only $O(N)$ operations and $O(N)$ storage.

4.3.2. *Solutions of linearized equations by improper integrals.* In this section we write the solutions of (25) as improper (but fast converging) integrals. These representations of solutions are convenient when the functions are discretized using splines or collocation methods, like in the cases where $X$ is known only at some points obtained experimentally.

**Lemma 4.2.** *Consider given* $\tilde{E} : \mathbb{T} \times [-1, 1] \rightarrow \mathbb{R}^2$ *which is a* $C^r$ *function* $r \in \mathbb{N} \cup \{\infty, \omega\}$, $r \geq 1$. *Assume that we are in the stable case* $(\lambda < 0)$.

*If* $\int_0^1 \tilde{E}(\theta, 0) \, d\theta = 0$, *the equation* (26) *has the solutions*
(30)

$$u(\theta, s) \equiv \alpha + \frac{1}{\omega} \int_0^\theta \tilde{E}(\sigma, 0) \, d\sigma + \int_0^\infty [\tilde{E}(\theta + \omega t, s e^{\lambda t}) - \tilde{E}(\theta + \omega t, 0)] dt.$$

*Furthermore, if* $\int \tilde{E}(\theta, 0) \neq 0$ *there is no* $C^0$ *solution of* (26). *The only solutions of* (26) *in* $C^0(\mathbb{T} \times [-1, 1])$ *are* (30).

*If* $r \geq 2$, $\int_0^1 \partial_s \tilde{E}(\theta, 0) d\theta = 0$ *the equation* (28) *has the solutions*

$$
\begin{aligned}
(31) \qquad u(\theta, s) = A(\theta) + s B(\theta) + \int_0^\infty e^{-\lambda t} [\tilde{E}(\theta + \omega t, s e^{\lambda t}) \\
- \tilde{E}(\theta + \omega t, 0) - s e^{\lambda t} \partial_s \tilde{E}(\theta + \omega t, 0)] \, dt,
\end{aligned}
$$

*where*

$$A(\theta) = \int_0^\infty e^{\lambda t} \tilde{E}(\theta - \omega t, 0) \, dt$$

$$B(\theta) = \alpha + \frac{1}{\omega} \int_0^\theta \partial_s \tilde{E}(\sigma, 0) \, d\sigma$$

*Furthermore, if* $\int_0^1 \partial_s \tilde{E}(\theta, 0) d\theta \neq 0$, *there is no* $C^1$ *solution of* (28).

*The only solutions of* (28) *in* $C^2(\mathbb{T} \times [-1, 1])$ *are those given by* (31).

*Proof.* We observe that if we particularize (26) to $s = 0$, we obtain

$$\omega \partial_\theta u(\theta, 0) = \tilde{E}(\theta, 0).$$

Hence, by the fundamental theorem of calculus, the only continuous solutions of (26) should satisfy

$$(32) \qquad u(\theta + \omega T, 0) - u(\theta, 0) = \int_0^T \tilde{E}(\theta + \omega t, 0) \, dt.$$

We see that the expression (32) is periodic in $T$ if and only if $\int_0^1 \tilde{E}(\theta, 0) \, d\theta = 0$. Hence, if $\int_0^1 \tilde{E}(\theta, 0) \, d\theta \neq 0$ there is no solution $u$. In the rest of the discussion we will assume $\int_0^1 \tilde{E}(\theta, 0) \, d\theta = 0$. We also note that the expression (32) is a solution of (26) on the set $s = 0$.

Also by the fundamental theorem of calculus and adding and subtracting terms we obtain that any solution of (26) should satisfy

(33)
$$u(\theta + \omega T, se^{\lambda T}) - u(\theta, s)$$
$$= \int_0^T [\tilde{E}(\theta + \omega t, se^{\lambda t}) - \tilde{E}(\theta + \omega t, 0)]\, dt + \int_0^T \tilde{E}(\theta + \omega t, 0)\, dt$$
$$= \int_0^T [\tilde{E}(\theta + \omega t, se^{\lambda t}) - \tilde{E}(\theta + \omega t, 0)]\, dt + u(\theta + \omega T, 0) - u(\theta, 0).$$

Because $u(\theta + \omega T, se^{\lambda t}) - u(\theta + \omega T, 0)$ converges to 0 as $T \to \infty$, and

$$|\tilde{E}(\theta + \omega t, se^{\lambda t}) - \tilde{E}(\theta + \omega t, 0)| \leq ce^{\lambda t}.$$

We obtain that the integral (33) is uniformly convergent and we obtain that the function $u$ is the only candidate for a solution.

Taking $T \to \infty$ in the expression (33) we have

$$u(\theta, s) = u(\theta, 0) - \int_0^\infty [\tilde{E}(\theta + \omega t, se^{\lambda t}) - \tilde{E}(\theta + \omega t, 0)]\, dt$$

and using again the fundamental theorem of calculus we have

$$u(\theta, 0) = u(0, 0) + \int_0^\theta \partial_\theta u(\sigma, 0)d\sigma = \alpha + \frac{1}{\omega} \int_0^\theta \tilde{E}(\sigma, 0)d\sigma,$$

which gives (30).

Since the integrand converges fast enough, we can compute the derivatives of the integral by computing the derivatives of the integrand and, therefore $u$ is indeed a solution.

To prove the second claim of Lemma 4.2, we proceed as before. We start by computing candidates for $u(\theta, 0)$ and $\partial_s u(\theta, 0)$ and then, we show that the integrand converges fast enough that we can justify that they are solutions. This strategy is very common in linearization problems and in invariant manifold theorems.

We observe that using the integrating factor $e^{-\lambda\theta/\omega}$ in equation (28) we have

$$-\lambda e^{-\lambda\theta/\omega} u(\theta, s) + e^{-\lambda\theta/\omega}(\omega\partial_\theta + \lambda s\partial_s)u(\theta, s) = e^{-\lambda\theta/\omega}\tilde{E}(\theta, s)$$

and therefore

$$(\omega\partial_\theta + \lambda s\partial_s)[e^{-\lambda\theta/\omega} u(\theta, s)] = e^{-\lambda\theta/\omega}\tilde{E}(\theta, s).$$

Using the fundamental theorem of calculus

$$e^{-\lambda(\theta+\omega T)/\omega}u(\theta+\omega T, se^{\lambda T}) - e^{-\lambda\theta/\omega}u(\theta, s) = \int_0^T e^{-\lambda(\theta+\omega t)/\omega}\tilde{E}(\theta+\omega t, se^{\lambda t})dt$$

and multiplying by $e^{\lambda\theta/\omega}$ we have the variation of parameters formula

$$(34) \qquad e^{-\lambda T}u(\theta + \omega T, se^{\lambda T}) - u(\theta, s) = \int_0^T e^{-\lambda t}\tilde{E}(\theta + \omega t, se^{\lambda t})dt.$$

We observe that using the variation of parameters formula (34) for $s = 0$, we obtain

$$e^{-\lambda T}u(\theta + \omega T, 0) - u(\theta, 0) = \int_0^T e^{-\lambda t}\tilde{E}(\theta + \omega t, 0)\, dt,$$

which multiplying by $e^{\lambda T}$ and performing the change of variables $\tilde{\theta} = \theta + \omega T$, becomes

$$u(\tilde{\theta}, 0) = u(\tilde{\theta} - \omega T, 0)e^{\lambda T} + \int_0^T e^{\lambda(T-t)}\tilde{E}(\tilde{\theta} - \omega(T - t), 0)\, dt$$

$$= u(\tilde{\theta} - \omega T, 0)e^{\lambda T} + \int_0^T e^{\lambda t}\tilde{E}(\tilde{\theta} - \omega t, 0)\, dt.$$

Taking limits as $T \to \infty$, we obtain the expression for $A$ in (31).

Now, we observe that if we take derivatives with respect to $s$ of (28) and evaluate at $s = 0$, we obtain

$$-\lambda\partial_s u(\theta, 0) + \omega\partial_\theta\partial_s u(\theta, 0) + \lambda\partial_s u(\theta, 0) = \partial_s\tilde{E}(\theta, 0),$$

which we rewrite as

$$\omega\partial_\theta[\partial_s u(\theta, 0)] = \partial_s\tilde{E}(\theta, 0).$$

Therefore, we have that if $\int \partial_s\tilde{E}(\theta, 0)\, d\theta = 0$, there is no periodic solution. Otherwise, we obtain $B$ in (31).

Equation (28) is obviously linear and we have found solutions $A(\theta)$, $sB(\theta)$ corresponding to RHS $\tilde{E}(\theta, 0)$, $s\partial_s\tilde{E}(\theta, 0)$. Therefore it suffices to find solutions for a RHS of (28) as

$$\tilde{\tilde{E}}(\theta, s) = \tilde{E}(\theta, s) - \tilde{E}(\theta, 0) - s\partial_s\tilde{E}(\theta, 0).$$

Again, we will find a candidate $\tilde{u}(\theta, s)$ and verify that indeed it is a solution.

The variation of parameters formula (34) gives

$$e^{-\lambda T}\tilde{u}(\theta + \omega T, se^{\lambda T}) - \tilde{u}(\theta, s) = \int_0^T e^{-\lambda t}\tilde{\tilde{E}}(\theta + \omega t, se^{\lambda t})\, dt.$$

We note that, because $\tilde{\tilde{E}}$ is $\mathcal{C}^2$ and $\tilde{\tilde{E}}(\theta, 0) = 0$, $\partial_s\tilde{\tilde{E}}(\theta, 0) = 0$

$$|\tilde{\tilde{E}}(\theta + \omega t, se^{\lambda t})| \le Ce^{2\lambda t}.$$

Hence, the integral in the RHS is convergent if we take the limit $T \to \infty$.

Since we have found the linear parts for $u(\theta, s)$, it is natural to guess that $\tilde{u}(\theta, s) \leq Cs^2$. Thus, we guess that the only solution for (28) is

$$(35) \qquad \tilde{u}(\theta, s) = \int_0^\infty e^{-\lambda t} \tilde{\tilde{E}}(\theta + \omega t, se^{\lambda t}) \, dt.$$

To prove that (35) is indeed a solution of (28), we note that taking derivatives under the integral sign (which is justified by the rapid convergence to zero of the integrand and its derivatives), we obtain that (35) satisfies (28).

Now we observe that, using the definition of $\tilde{u}$, if there was a solution $u$ of (28), it would satisfy

$$-\lambda(u - \tilde{u}) + (\omega \partial_\theta + \lambda s \partial_s)(u - \tilde{u}) = \tilde{E}(\theta, 0) + s\partial_s \tilde{E}(\theta, 0),$$

but we have already established the uniqueness of this previous approximation. $\qquad \square$

Of course, from the form of the solutions (30) and (31) we can also obtain regularity properties, but this will be done in Section 5.

Note that, since $\lambda < 0$, the integrals defining $A$ are uniformly convergent. The integrals in (31) are also uniformly convergent because, by Taylor's theorem,

$$|\tilde{E}(\theta + \omega t, se^{\lambda t}) - \tilde{E}(\theta + \omega t, 0) - se^{\lambda t} \partial_s \tilde{E}(\theta + \omega t, 0)]| \leq Ce^{2\lambda t}.$$

The representation of the solutions by convergent integrals is rather convenient if we represent the functions by the values on a grid and then, interpolate using splines.

If we fix the values of $\theta$ and $s$, the evaluation of the integrals (30) and (31) can be done in a number of operations which depends only on the accuracy required. It should be moderate because the integrands converge rather rapidly. Hence, we expect that the number of operations to compute $N$ discretization points will be $\mathcal{O}(N)$, but the constant could be large. On the other hand, the evaluation of the integrals at different points is clearly paralellizable and the discretization allows to use more points in the places where the function is worse behaved. Note also that it is possible to use a mixed representation. Using splines in $\theta$ and using Taylor series in $s$. See Section 7 for a more detailed discussion.

4.4. **Algorithm for the quasi-Newton step.** In this section we specify step by step the implementation of the Quasi-Newton step. We call attention to the fact that the ingredients of the Quasi-Newton step consist only of algebraic operations, composition of functions, taking derivatives and solving cohomology equations.

The theory of solvability of linearized equations developed in Lemma 4.1 indicates how to approach the solution of (25). We determine the unknowns $\sigma, \eta$ so that the equations are all solvable, namely

(36)
$$\sigma = \int_0^1 \tilde{E}_1(\theta, 0)d\theta,$$
$$\eta = \int_0^1 \partial_s \tilde{E}_2(\theta, s)_{|s=0}d\theta.$$

Therefore, we are led to the following algorithm

**Algorithm 4.3.** *Consider a vector field* $X : \mathbb{R}^2 \to \mathbb{R}^2$. *Given* $K : \mathbb{T} \times [-1, 1] \to \mathbb{R}^2$, $\omega \in \mathbb{R}$, $\lambda \in \mathbb{R}$. *Compute:*

(1) $\alpha \leftarrow X \circ K$
(2) $\beta \leftarrow DK$
(3) $E \leftarrow \alpha - \beta\, A_{\omega\lambda}$
(4) $\tilde{E} = DKE$. *Denote* $\tilde{E} = (\tilde{E}_1, \tilde{E}_2)$.
(5) $\sigma = \int_0^1 \tilde{E}_1(\theta, 0)\, d\theta$
    $\eta = \int_0^1 \partial_s \tilde{E}_2(\theta, s)_{|s=0}\, d\theta$
(6) *Find* $W_1$ *solving*

$$(\omega\partial_\theta + \lambda s \partial_s)W_1 = \tilde{E}_1 - \sigma.$$

*We also impose the normalization* $\int W_{1|s=0} = 0$, *so that the solution is unique*
(7) *Find* $W_2$ *solving*

$$(\omega\partial_\theta + \lambda s \partial_s)W_2 - \lambda W_2 = \tilde{E}_2 - \eta s$$

*We also impose the normalization* $\int \partial_s W_{2|s=0} = 0$, *so that the solution is unique*
(8) *Denote* $W = (W_1, W_2)$. *The improved solution is*

$$\tilde{K} = K + DKW, \quad \omega = \omega + \sigma, \quad \lambda = \lambda + \eta$$

One remarkable feature of Algorithm 4.3 is that even if it is a quadratically convergent algorithm, at no stage of the algorithm, it requires to compute (much less to invert) a matrix of the dimension of the discretization. We only need to perform algebraic operations among functions, computing derivatives and solving cohomology equations.

All the above operations can be implemented either in Taylor-Fourier series or using a discretization in a grid of points and interpolate using e.g. splines. Note that, as indicated in Section 4.3.1, the solution of the linearized equations is diagonal in Fourier-Taylor coefficients.

If we discretize the functions with $N$ points, the storage required is $\mathcal{O}(N)$ and the number of operations is $\mathcal{O}(N \log(N))$ if we use the FFT

to switch between representations or $\mathcal{O}(N)$ – with a larger constant – if we use the real space representation in a grid of points and evaluate at intermediate points using (one-dimensional) splines.

*Remark* 4.4. Notice that the solution $W_1$ obtained in step (6) is unique up to the addition of a constant and $W_2$ computed in step (7) is unique up to the adition of a multiple of the first order coefficient.

The solutions obtained for $\eta, \sigma$ are unique.

The undeterminacy in the solutions of (6) and (7) can be used to achieve other normalizations.

We also emphasize that, since the algorithm was obtained using only algebraic or calculus identities from the linearized equation, these undeterminacies are the only undeterminacies of the modified Newton equation. The uniqueness of the linearized equation will be the basis of the proof of the uniqueness result in Theorem 6.1.

*Remark* 4.5. It is remarkable to note that the computation of the limit cycles and the isochrones, requires less computational effort than the computation of the limit cycle alone.

If we were to compute the parameterization of the limit cycle alone by a Newton method, we would need to invert a full matrix of derivatives. Computing at the same time the limit cycle and the first order of the isochrones, one can get a fast algorithm. The computation of the higher order of the isochrones does not change the leading order of the requirements in storage or in the operation count.

Similar phenomena have appeared other times in computational dynamics. Sometimes it is advantageous to consider the linearized equations and "reduce" them. In [JS92, HL06], the reduction of the linearized equations required some additional computation. In our case, the reducibility is automatic.

## 5. Convergence of the iterative step and Proof of Theorem 3.2

In this section we prove Theorem 3.2 that establishes the convergence of Algorithm 4.3 provided that we start with a sufficiently approximate solution of (4).

Note that Theorem 3.2 has the format of *a-posteriori* results of numerical analysis. We show that if the initial approximation solves the equation with sufficient accuracy depending on explicit "condition numbers", then there is a true solution and we can bound the difference between the initial approximation and the true solution by the residual of (4) evaluated on the initial approximation.

The *a-posteriori* results such as Theorem 3.2 are very typical of results on convergence of methods based in Newton-Kantorovich algorithms, which have an iterative method leading to a fixed point. Nevertheless, since our iterative step involves taking derivatives we will have to use Nash-Moser estimates rather than the more elementary Kantorovich ones. This will require introducing norms to measure the distance between functions, etc. Since we are using a Nash-Moser method, we will need to use families of norms. Nash-Moser methods are very robust and can work with several norms. For simplicity, we will only discuss the most customary supremum norms introduced in Definition 3.1.

5.1. **Some elementary properties of the norms in Definition 3.1.** In this section we review some elementary properties of the norms introduced in Section 3.1

**Proposition 5.1.** *Consider the notation introduced in Definition 3.1. For any $f, g \in \mathcal{A}_\rho$ and $K, L \in \mathcal{A}_{\beta,\rho}$, we have*

$$(37) \qquad \begin{aligned} \|f \cdot g\|_\rho &\leq \|f\|_\rho \|g\|_\rho \\ \|K \cdot L\|_{\beta,\rho} &\leq \|K\|_{\beta,\rho} \|L\|_{\beta,\rho} \end{aligned}$$

The proof of Proposition 5.1 is immediate if we just observe that the supremum of the product is less than the product of the supremums.

**Proposition 5.2.** *For any $\delta > 0$, we have for any $f \in \mathcal{A}_\rho$, $K \in \mathcal{A}_{\beta,\rho}$.*

$$(38) \qquad \begin{aligned} \|\partial_\theta f\|_{\rho-\delta} &\leq C\delta^{-1} \|f\|_\rho \\ \|\partial_s K\|_{\beta-\delta,\rho} &\leq C\delta^{-1} \|K\|_{\beta,\rho} \\ \|\partial_\theta K\|_{\beta,\rho-\delta} &\leq C\delta^{-1} \|K\|_{\beta,\rho} \end{aligned}$$

This is a very standard result in complex analysis that follows from Cauchy formula for the derivative as a contour integral. See [Ahl78, SZ65].

**Proposition 5.3.** *Let $X$ be an analytic vector field in a domain $\mathcal{C} \subset \mathbb{C}^2$. Let $K : \mathbb{T}_\rho \times B_\beta \to \mathbb{C}^2$ be such that*

$$\mathrm{dist}(K(\mathbb{T}_\rho \times B_\beta), \mathbb{C}^2 - \mathcal{C}) \geq \zeta > 0.$$

*Then,*

- $X \circ K \in \mathcal{A}_{\beta,\rho}$. *In particular, $X \circ K$ is analytic on $\mathbb{T}_\rho \times B_\beta$.*
- *For all $\gamma : \mathbb{T}_\rho \times B_\beta \to \mathbb{C}^2$ with $\|\gamma\|_{\beta,\rho}$ sufficiently small, we have:*

$$(39) \qquad \|X \circ (K+\gamma) - X \circ K - DX \circ K\, \gamma\|_{\beta,\rho} \leq C\|\gamma\|_{\beta,\rho}^2$$

The proof of Proposition 5.3 follows from the observation that, for each $x \in \mathbb{T}_\rho$ we can use the Taylor's theorem and then take the supremum. This gives that the constant $C$ apppearing in (39) can be taken to be just

$$C = \frac{1}{2} \sup_{x \in \mathcal{C}} ||D^2 X||.$$

Another useful property of the norms is that they are log convex in $\rho$. This is just *Hadamard's three circle theorem* [Ahl78, SZ65]. For all $\rho_1, \rho_2 > 0$, $0 \leq \alpha \leq 1$, we have

$$(40) \qquad ||\phi||_{\alpha\rho_1 + (1-\alpha)\rho_2} \leq ||\phi||_{\rho_1}^\alpha ||\phi||_{\rho_2}^{1-\alpha}.$$

5.2. **Estimates for the iterative step.** In this section we quantify the argument presented heuristically showing that indeed the error in (4) after the iterative step is bounded by the square of the error before the step. There are some subtleties (standard in KAM theory) that need to be taken into account: a) The bounds after the step are in a slightly smaller domain; b) The bounds have constants that blow up (like a power) on the loss of analyticity; c) The bounds have constants that depend on some non-degeneracy conditions which can be written explicitly and consist of algebraic expressions involving derivatives of $K$.

**Lemma 5.4.** *Assume that $X$ is analytic in some domain $U \subset \mathbb{C}^2$. Let $K : \mathcal{D}_{\beta,\rho} = \mathbb{T}_\rho \times B_\beta \to U$ belong to $\mathcal{A}_{\beta,\rho}$.*
*Assume that*

$$d(\mathrm{Range}(K(\mathcal{D}_{\beta,\rho}), \mathbb{C}^2 - U) \geq \zeta > 0.$$

*Assume furthermore that for some $m \geq 0$,*

$$(41) \qquad \begin{aligned} &\|K\|_{\beta,\rho} \leq m, \\ &\|DK\|_{\beta,\rho}, \|D^2K\|_{\beta,\rho} \leq m_+, \\ &\|DK^{-1}\|_{\beta,\rho} \leq m_- \,, \\ &\omega \geq \tilde{m}\,, \text{ and } \lambda \leq -\tilde{m}. \end{aligned}$$

*Let $E$ be the error function defined as*

$$E = X \circ K - DK A_{\omega,\lambda},$$

*and $\delta > 0$ be such that*

$$(42) \qquad \delta^{-1} m \|E\|_{\beta,\rho} \leq \zeta/100.$$

*Then, there is a constant $C$ depending only on $\zeta, m_+, m_-, \tilde{m}$ such that, the improved solution $(K + \Delta, \omega + \sigma, \lambda + \eta)$ obtained after the Quasi-Newton step specified in Algorithm 4.3 satisfies:*

$$(43) \quad \|X \circ (K + \Delta) - D(K + \Delta)A_{\omega+\sigma,\lambda+\eta}\|_{\beta-\delta,\rho-\delta} \leq C\delta^{-1}\|E\|_{\beta,\rho}^2$$

*Proof.* The proof follows just walking through the argument presented in Section 4.4 and adding and subtracting appropriate terms in the linear expansion.

Our goal is to estimate the error of the improved approximation $X + \Delta$, $\omega + \sigma$, $\lambda + \eta$ where $\Delta = DKW$, and $W, \sigma, \eta$ are obtained through Algorithm 4.3 (see points (4)-(7)).

Remember that the prescription to compute the Fourier coefficients of the function $W$ is specified in (27) and (29). Then, using that $\tilde{E} = DKE$ and Proposition 5.1, we have that

$$\begin{aligned}
|\sigma|, |\eta| &\leq \|\tilde{E}\|_{1,\rho} \\
&\leq \|DK\|_{1,\rho}\|E\|_{1,\rho} \\
&\leq C\|E\|_{1,\rho}
\end{aligned}$$

$$(44)$$

$$\|W\|_{1,\rho} \leq C\left(\frac{1}{\omega} + \frac{1}{\lambda}\right)\|E\|_{1,\rho}$$

where $C$ is a constant that depends on $m_+$.

The following identity is obtained just adding and subtracting some terms (the ones we declared as the leading coefficients and the ones that we cancel) and grouping:

$$\begin{aligned}
X \circ &(K + \Delta) - D(K + \Delta)A_{\omega+\sigma,\,\lambda+\eta} \\
&= X \circ (K + DKW) - D(K + DKW)(A_{\omega,\lambda} + A_{\sigma,\eta}) \\
&= X \circ (K + DKW) - DKA_{\omega,\lambda} - DKA_{\sigma,\eta} - D^2KWA_{\omega\lambda} \\
&\quad - D^2KWA_{\sigma,\eta} - DKDWA_{\omega,\lambda} - DKDWA_{\sigma,\eta} \\
&= X \circ (K + DKW) - X \circ K - DX \circ KDKW \\
&\quad + [DX \circ KDK - D^2KA_{\omega,\lambda} - DKDA_{\omega,\lambda}]W \\
&\quad + X \circ K - DKA_{\omega,\lambda} \\
&\quad + DKDA_{\omega,\lambda}W - DKDWA_{\omega,\lambda} - DKA_{\sigma,\eta} \\
&\quad - DKDWA_{\sigma,\eta} \\
&\quad - D^2KWA_{\sigma,\eta}.
\end{aligned}$$

$$(45)$$

The different lines (which we denote $\ell_1 - \ell_6$) in the last expression of (45) can be estimated as follows.

The third line ($\ell_3$) in (45) is just $E$ and we observe that the point of the solution of the quasi-Newton method (23) is that $W, \sigma, \eta$ are chosen so that the third and fourth line of (45) cancel exactly.

We recall that we denote by $C$ numbers that are controlled by some function of the condition numbers $m$ in (41) as well as $\omega, \lambda$ and $\sup_{x \in \mathcal{C}} |D^2 X|$.

The first line of (45) can be estimated using the Taylor remainder and the bound for $\|W\|_{\beta,\rho}$ obtained in (44). We obtain

$$
(46) \quad
\begin{aligned}
\|\ell_1\|_{\beta,\rho} &\leq \frac{1}{2}\|D^2 X\|_{\mathcal{C}} \left(\|DK\|_{\beta,\rho}\|W\|_{\beta,\rho}\right)^2 \leq \frac{1}{2}C\|W\|_{\beta,\rho}^2 \\
&\leq C\|E\|_{\beta,\rho}^2.
\end{aligned}
$$

The second line of (45) can be estimated observing that the term in parenthesis is the derivative of $E$ (see equation (19)). Then, using Cauchy bounds (Proposition 5.2) and the bound for $\|W\|_{\beta,\rho}$ obtained in (44) as well as the Banach algebra properties (Proposition 5.1), we obtain

$$
\begin{aligned}
\|\ell_2\|_{\beta-\delta,\rho-\delta} &\leq C\delta^{-1}\|E\|_{\beta,\rho}\|W\|_{\beta,\rho} \\
&\leq C\delta^{-1}\|E\|_{\beta,\rho}^2.
\end{aligned}
$$

The fifth and sixth line of (45) can be estimated straightforwardly using the estimates for $\sigma, \eta$ and $W$ in (44) and Cauchy bounds (Proposition 5.2) by

$$
(47) \quad
\begin{aligned}
\|\ell_5\|_{\beta-\delta,\rho-\delta} &\leq C\delta^{-1}\|W\|_{\beta,\rho}|(\sigma,\eta)| \\
&\leq C\delta^{-1}\|E\|_{\beta,\rho}^2 \\
\|\ell_6\|_{\beta,\rho} &\leq C\|W\|_{\beta,\rho}|(\sigma,\eta)| \\
&\leq C\|E\|_{\beta,\rho}^2
\end{aligned}
$$

Taking the minimum of these estimates we obtain the bound in (43). $\qquad\square$

*Remark* 5.5. In contrast with the usual KAM problems, we do not lose derivatives in the solutions of the linearized problem. In that case, the lost of derivatives in the Newton step is because the functional equation (4) involves taking derivatives to straighten the vector field and composition of functions (which is not a differentiable operator unless one loses some domain).

5.3. **Repeating the iteration and end of the proof of Theorem 3.2.** The proof is very standard in KAM theory. We follow very closely the presentation in [Lla01] for KAM problems.

We assume by induction that the iterative step can be carried out $n$ times (i.e., that hypothesis (42) is verified for the first $n$ steps). We denote by subindices $n$ the objects after $n$ steps of the iterative process. We will show that, under certain assumptions on the size of $\delta_0$ and the error in the intial approximation $\varepsilon_0$, which will be independent of $n$, hypothesis (42) will be verified for $n + 1$. Moreover, we will show that the error for successive approximations decreases very fast (superexponentially).

We start by fixing

$$(48) \qquad\qquad \delta_n = \frac{1}{4}\delta_0 2^{-n}.$$

We will show that this choice of $\delta_n$ is acceptable when the error in the initial approximation $\varepsilon_0$ is small enough.

The condition numbers $m_-, m_+, \tilde{m}$ will be changing in the iteration. We will assume inductively that they are twice as bad as the initial value. We will show that this induction assumption is maintained if $\varepsilon_0$ is small enough. We will denote by $C$ the constant that corresponds to the values of $m, m_+$ and $\tilde{m}$, which are twice the original values.

Denoting by $\varepsilon_n$ the value of the error at step $n$, the estimates for the iterative step, can be written

$$(49) \qquad\qquad \varepsilon_n \leq C(\delta_0 2^{-n-1})^{-1}\varepsilon_{n-1}^2.$$

Repeating (49), we obtain

$$
\begin{aligned}
(50) \qquad \varepsilon_n &\leq C\delta_0^{-1}2^{n+1}\varepsilon_{n-1}^2 \\
&\leq (C\delta_0^{-1})2^{n+1}(C\delta_0^{-1})^2 2^{2n}\varepsilon_{n-2}^{2\cdot 2} \\
&\leq (C\delta_0^{-1})^{1+2+2^2+\cdots+2^{n-1}}2^{(n+1)+2(n)+2^2(n-1)+\cdots+2^n}\varepsilon_0^{2^n} \\
&\leq (C\delta_0^{-1})^{2^n}2^{2^{n+1}\sum_{k=1}^{n+1}k2^{-k}}\varepsilon_0^{2^n} \\
&\leq (C\delta_0^{-1})^{2^n}2^{2^{n+1}\sum_{k=0}^{\infty}k2^{-k}}\varepsilon_0^{2^n} \\
&\leq (C\delta_0^{-1})^{2^n}2^{2^{n+2}}\varepsilon_0^{2^n} \\
&\leq (C\delta_0^{-1}2^2\varepsilon_0)^{2^n}
\end{aligned}
$$

Notice that if $C\delta_0^{-1}\varepsilon_0 < 1$, then $\varepsilon_n$ is superexponentially small. We will now show that the fact that $\varepsilon_n$ decresases superexponentially while $\delta_n$ decreases only exponentially (48) has the consequence that the inductive assumption (42) will be satisfied for all the iterative steps if the initial error $\varepsilon_0$ is small enough.

Furthermore, denoting $\mathcal{D}_{\beta_n,\rho_n}$ the domain of definition of $K_n$ with $(\beta_n, \rho_n) = (\beta_{n-1} - \delta_{n-1}, \rho_{n-1} - \delta_{n-1})$, we have that

$$\|DK_n - DK_0\|_{\eta_n,\rho_n} \leq \sum_{j=1}^{n} \|DK_j - DK_{j-1}\|_{\eta_j,\rho_j} \leq \sum_{j=1}^{n} C\varepsilon_j \delta_j^{-1}$$

$$\leq \sum_{j=1}^{n} (C\delta_0^{-1}\varepsilon_0)^{2^j} \delta_0^{-1} 2^j$$

Similarly, we obtain

$$\|D^2 K_n - D^2 K_0\|_{\eta_n,\rho_n} \leq \sum_{j=1}^{n} (C\delta_0^{-1}\varepsilon_0)^{2^j} \delta_0^{-2} 2^{2j}$$

Hence, we see that if $\varepsilon_0$ is small enough we obtain that the assumption that the change of $m_+$ is small enough is satisfied.

Similarly, we see that the other smallness of the change are satisfied if $\varepsilon_0$ is sufficiently small. We get, therefore that the inductive assumptions amount to a finite number of smallness assumptions on $\varepsilon_0$.

Note also that, adding and subtracting terms and using (44), we have

(51)
$$\|K_0 - K_\infty\|_{\beta_\infty,\rho_\infty} \leq \sum_{n=0}^{\infty} \|K_n - K_{n+1}\|_{\beta_\infty,\rho_\infty}$$
$$\leq \sum_{n=0}^{\infty} \|K_n - K_{n+1}\|_{\beta_{n+1},\rho_{n+1}}$$
$$\leq \sum_{n=0}^{\infty} C\varepsilon_n \leq C\varepsilon_0$$

The last inequality, of course, depends on $\varepsilon_0\delta^{-1}$ being sufficiently small so that the superexponential convergence implies that the dominant term in the infinite sum above is the first one.

Observe that the recurrence for the error (49) can be rewritten more transparently as

$$\varepsilon_n \leq \left(C(\delta_0 2^{-n-1})^{-1}\varepsilon_{n-1}\right) \varepsilon_{n-1}$$

Using (50) we obtain that for $\varepsilon_0$ sufficently small, we have $C(\delta_0 2^{-n-1})^{-1}\varepsilon_{n-1} \leq 1/2$. Hence, we can estimate the sums by a geometric series with an initial term $\varepsilon_0$ and ratio $1/2$.

Analogous consideration lead to estimates of $|\omega - \omega^*|$, $|\lambda - \lambda^*|$.

*Remark* 5.6. For the experts in KAM theory, we note that in our case, the size of the correction is bounded by the error without any factor

from the loss of analyticity. The factors in $\delta^{-1}$, come only from the fact that the functional we are studying is not differentiable. In the regular KAM theory, the corrections at the step require a factor of the loss of differentiability.

In both cases, we obtain that the total change in the function is bounded by a multiple of the first step. In the standard KAM case, this first step is the error times a power of the analyticity loss. In our case, the step is bounded by a step of the error.

## 6. Some remarks and extensions of the analytic proof

### 6.1. Local uniqueness.

**Theorem 6.1.** *Let* $(K_1, \omega_1, \lambda_1)$ *and* $(K_2, \omega_2, \lambda_2)$ *be solutions of* (4) *for the same vector field* $X$. *If*

$$(52) \qquad \|K_1 - K_2\|_{\beta,\rho}, |\omega_1 - \omega_2|, |\lambda_1 - \lambda_2| \leq C$$

*for a constant* $C$ *that depends on the condition numbers of the solution* $(K_1, \omega_1, \lambda_1)$ *and* $\beta, \rho$, *then*

$$(53) \qquad \omega_1 = \omega_2, \qquad \lambda_1 = \lambda_2,$$

*and there is* $\theta_0, b \in \mathbb{R}$, *such that*

$$K_2(\theta, s) = K_1(\theta + \theta_0, bs).$$

Note that, as anticipated in Remark 2.2, the only non-uniqueness in (4) is the choice of origin in phase and the choice of scale for the parameter on the isochrones. Of course, the isochrones and the limit cycle are not affected by these choices of the parameterization.

6.1.1. *Proof of Theorem 6.1.* The proof is very similar to the local uniqueness results in other papers [LGJV05, CdlL10] which also use automatic reducibility methods. The key observation is that the linearized equation admits a unique solution if we impose a normalization. In the language of abstract implicit function theorems, this is expressed as saying that the linearized equation admits a "left inverse". For a discussion of this from an abstract point of view, we refer to [CdlL10, Appendix A].

To overcome the ambiguity pointed out in Remark 2.2, we need to introduce a definition of normalized solutions.

**Definition 6.2.** *Given a solution $(K_1, \omega_1, \lambda_1)$ of (4), we say that another embedding $K_2$ is $K_1$-normalized when*

(54)
$$\int d\theta \, \Pi_1 \left[ (K_2 - K_1) DK_1^{-1} \right]_{s=0} = 0,$$
$$\int d\theta \, \Pi_2 \left[ DK_2 DK_1^{-1} \right]_{s=0} = 1,$$

*where $\Pi_1, \Pi_2$ denote the projections over the components.*

The interpretation of (54) is that, when we express the difference between the solutions in the natural frame of reference of $K_1$, then the first coordinate has average zero. Furthermore, that the vector field representing the stable directions of the solutions have integral 1. These normalizations are natural since they eliminate the indetermimations we already identified in the solutions of (4), namely the change of the origin in the angle variable and the change of scale in the linear variable.

**Lemma 6.3.** *There exists a constant $C$ such that if $\|K_1 - K_2\|_{\beta, \rho} \leq C$, then there exist small $\theta_0$ and $b$ close to 1 such that if we denote $B_{\theta_0, b}(\theta, s) = (\theta + \theta_0, bs)$, then $K_2 \circ B_{\theta_0, b}$ is $K_1$-normalized.*

The proof is similar to Lemma 14 in [LGJV05]. It is based on the application of the implicit function theorem to the function $F^K(\theta_0, b)$ defined by evaluating (54) at $K \circ B_{\theta_0, b}$. That is,

(55)
$$F_1^K(\theta_0, b) = \int d\theta \, \Pi_1 \left[ (K \circ B_{\theta_0, b} - K_1) DK_1^{-1} \right]_{s=0}$$
$$F_2^K(\theta_0, b) = \int d\theta \, \Pi_2 \left[ DK \circ B_{\theta_0, b} DK_1^{-1} \right]_{s=0}.$$

Hence, to prove Theorem 6.1, it suffices to show that if we have two solutions of (4), $(K_1, \omega_1, \lambda_1)$ and $(K_2, \omega_2, \lambda_2)$ and that $K_2$ is $K_1$-normalized, then we have that they are equal. To do so and, it will be useful to introduce a more abstract point of view similar to [CdlL10]. We introduce the notation

(56)
$$\mathcal{T}(K, \omega, \lambda) \equiv X \circ K - DK A_{\omega, \lambda}$$

so that (4) can be written as

(57)
$$\mathcal{T}(K, \omega, \lambda) = 0.$$

If we are given two solutions $(K_1, \omega_1, \lambda_1)$ and $(K_2, \omega_2, \lambda_2)$ of (57), we can write, using Taylor's theorem from one to the other,

(58)
$$
\begin{aligned}
0 &= \mathcal{T}(K_2, \omega_2, \lambda_2) \\
&= \mathcal{T}(K_1, \omega_1, \lambda_1) + D\mathcal{T}(K_1, \omega_1, \lambda_1)[K_2 - K_1, \omega_2 - \omega_1, \lambda_2 - \lambda_1] + R \\
&= D\mathcal{T}(K_1, \omega_1, \lambda_1)[K_2 - K_1, \omega_2 - \omega_1, \lambda_2 - \lambda_1] + R,
\end{aligned}
$$

where $R$ is the reminder of the Taylor expansion of the functional $\mathcal{T}$. The fact that $\mathcal{T}$ is differentiable and the form of the derivative have been established in Lemma 5.4.

Note that $\mathcal{T}$ involves only composing $X$ on the right with $K$, taking derivatives of $K$ and performing some algebraic operations. Hence, we have that

(59) $\qquad ||R||_{\rho-\delta} \leq C\delta^{-1}(||K_2 - K_1||_\rho^2 + |\omega_2 - \omega_1|^2 + |\lambda_2 - \lambda_1|^2).$

The identity (58) relates the increments in the unknown to the Taylor estimates in exactly the same way that the corrections of the Newton method were related to the error. We can regard (58) as an equation for $K_2 - K_1, \omega_2 - \omega_1, \lambda_2 - \lambda_1$. Using that these equations have unique solutions (because $K_2$ is $K_1$-normalized), we have that:

(60)
$$
\begin{aligned}
||K_2 - K_1||_{\rho-2\delta}, |\omega_2 - \omega_1|, |\lambda_2 - \lambda_1| &\leq C\delta^{-1}||R||_{\rho-\delta} \\
&\leq C\delta^{-2}(||K_2 - K_1||_\rho^2 + |\omega_2 - \omega_1|^2 + |\lambda_2 - \lambda_1|^2).
\end{aligned}
$$

Using Hadamard's three circle theorem (40), we obtain:

$$||K_2 - K_1||_\rho^2 \leq ||K_2 - K_1||_{\rho+2\delta}||K_2 - K_1||_{\rho-2\delta}.$$

Hence,

(61)
$$
\begin{aligned}
&||K_2 - K_1||_{\rho-2\delta} + |\omega_2 - \omega_1| + |\lambda_2 - \lambda_1| \\
&\quad \leq C\delta^{-2}(||K_2 - K_1||_{\rho+2\delta} + |\omega_2 - \omega_1| + |\lambda_2 - \lambda_1|) \cdot \\
&\quad\quad \cdot (||K_2 - K_1||_{\rho-2\delta} + |\omega_2 - \omega_1| + |\lambda_2 - \lambda_1|)
\end{aligned}
$$

Thefore if $C\delta^{-2}(||K_2-K_1||_{\rho+2\delta}+|\omega_2-\omega_1|+|\lambda_2-\lambda_1|) < 1$, we conclude that $K_1 = K_2$, $\omega_1 = \omega_2$, $\lambda_1 = \lambda_2$. The statement of Theorem 6.1 is obtained just by redefining $\rho$. $\qquad\square$

6.2. **Dependence on parameters.** In many applications, the models depend on extra parameters. We will show how the automatic reducibility methods used in the proof of Theorem 3.2 lead to very efficient computations of the perturbative expansions with respect to these parameters. We will also show, following [Mos67], that these perturbative expansions converge.

6.2.1. *Lipschitz dependence on parameters.* If we consider a family of vector fields $X_\mu$ and we assume Lipschitz dependence of the vector field with respect to the parameter $\mu$, we can use Theorem 3.2 to obtain Lipschitz dependence of the solution with respect to the parameter $\mu$.

Following the notation introduced in (56), denote

$$\mathcal{T}_\mu(K, \omega, \lambda) \equiv X_\mu \circ K - DK A_{\omega,\lambda},$$

so that a solution $(K_\mu, \omega_\mu, \lambda_\mu)$ of (4) for the vector field $X_\mu$ can be written as

(62) $$\mathcal{T}_\mu(K_\mu, \omega_\mu, \lambda_\mu) = 0.$$

Consider given $(K_\mu, \omega_\mu, \lambda_\mu)$ satisfying (62). Then, using that $X_\mu$ is Lipschitz dependent with respect to parameter $\mu$, we clearly have

$$\|\mathcal{T}_{\tilde{\mu}}(K_\mu, \omega_\mu, \lambda_\mu)\|_{\beta-\delta,\rho-\delta} = \|\mathcal{T}_{\tilde{\mu}}(K_\mu, \omega_\mu, \lambda_\mu) - \mathcal{T}_\mu(K_\mu, \omega_\mu, \lambda_\mu)\|_{\beta-\delta,\rho-\delta}$$
$$= \|X_{\tilde{\mu}} \circ K_\mu - X_\mu \circ K_\mu\|_{\beta-\delta,\rho-\delta} \leq C\delta^{-1}|\mu - \tilde{\mu}|.$$

So, we have that $(K_\mu, \omega_\mu, \lambda_\mu)$ is as an approximate solution of (62) for values of the parameter $\mu$ close to the original one.

Hence, we can apply Theorem 3.2 and obtain that for $|\mu - \tilde{\mu}|$ sufficiently small, there exists $K_{\tilde{\mu}}$ an analytic local diffeomorphism and $\omega_{\tilde{\mu}}, \lambda_{\tilde{\mu}} \in \mathbb{R}$, such that

$$\|K_\mu - K_{\tilde{\mu}}\|_{\beta-2\delta,\rho-2\delta}, |\omega_\mu - \omega_{\tilde{\mu}}|, |\lambda_\mu - \lambda_{\tilde{\mu}}| \leq C\delta^{-1}|\mu - \tilde{\mu}|.$$

6.2.2. *Expansions on parameters.* Consider given a parametric family of vector fields $X_\mu$ as well as a solution $(K_0, \omega_0, \lambda_0)$ of (62) for $\mu = 0$. We will first discuss how to compute a formal solution of (62) for $\mu \neq 0$, by considering asymptotic expansions on the parameter $\mu$,

(63)
$$K_\mu = \sum_n \mu^n K^n$$
$$\omega_\mu = \sum_n \mu^n \omega^n, \quad \lambda_\mu = \sum_n \mu^n \lambda^n.$$

We will show that Theorem 3.2 leads to efficient calculations of these asymptotic expansions, as well as its convergence.

*Efficient calculation of asymptotic expansions*

We will discuss two different methods to compute the asymptotic expansions efficiently. We will first discuss the order by order method. The other method is based on the philosophy of quasi-Newton methods. *Order by order method.* Inductively we assume that we have computed the expansion (63) up to order $n - 1$ and we want to show that it is possible to compute the expansion up to order $n$.

We substitute (63) into (4) and by matching coefficients of order $n$, we obtain

$$(64)\quad (DX_\mu \circ K_0)K^n - DK^n A_{\omega_0,\lambda_0} - DK_0 A_{\omega^n,\lambda^n} = S_n(K_0,\dots,K^{n-1}),$$

where $S_n$ is an explicit polynomial expression in $K^0,\dots,K^{n-1}$ whose coefficients are derivatives of $X_\mu$ evaluated at $K_0$. These coefficients can be calculated efficiently using the methods of automatic differentiation when $X_\mu$ is formed using elementary functions (polynomials, exponentials, trigonometric functions, etc).

We observe that equations (64) are identical to (18); the equations we studied in Section 4.1. Hence, we can use the same method used there with some minor differences that we discuss now.

Note that because $K_0$ satisfies exactly the invariance equation (4), the factorization of the equation (64) into elementary steps achieved in Algorithm 4.3 holds exactly. Furthermore, all the auxiliary quantities involved in the factorization need to be computed only once because for all steps we only consider the linearization around $K_0$, which does not change during the iteration.

*A quadratically convergent method for perturbative expansions.* An even faster method to compute the perturbative expansions consists in considering $K(\theta,s,\mu)$; that is $K$ is a function of the parameter $\mu$. It is easy to see that the Algorithm 4.3 lifts to functions of three variables and that one can also obtain quadratic convergence in the space of functions in these three variables using the argument in Section 3.

*Convergence of perturbative expansions*

Convergence of perturbative expansions is guaranteed by Theorem 3.2. It suffices to take the solution of (62) for $\mu = 0$ as an approximate solution of (62), and by Theorem 3.2 we have that there exists a solution $(K_\mu,\omega_\mu,\lambda_\mu)$ of (62) for $\mu$ small and complex. Then, using Lemma 6.3. we can assume that the solutions are $K_0$-normalized in the sense of Definition 6.2. We also know that functions $(K_\mu,\omega_\mu,\lambda_\mu)$ are differentiable for $\mu$ small and complex. Hence, they are analytic in $\mu$.

6.3. **Finite differentiablity.** There is a standard procedure in [Mos66b, Mos66a], systematized and extended in [Zeh75] that shows that one can deduce results for finite differentiable problems from quantitative results such as Theorem 3.2 for analytic problems.

The key is the following Lemma characterizing the finitely differentiable functions by the speed of approximation by analytic functions.

**Lemma 6.4.** *A function $f : \mathbb{T}^d \times B^l$ is $r$ times continuously differentiable, $r \in \mathbb{N}$ and the $r$ derivative is Hölder continuous with exponent*

$\alpha$, $0 < \alpha < 1$ *if and only if we can find a sequence of functions $f_n$, each of them analytic in a complex extension of size $\rho_n = 2^{-n}$ and such that*

- $||f_n - f_{n-1}||_{\rho_n} \leq C2^{-n(r+\alpha)}$
- $||f_n - f||_{C^0} \to 0$

A very streamlined poof of Lemma 6.4 is found in [Zeh75]. It is also well known that the characterization given by Lemma 6.4 is false for $\alpha = 0, 1$.

Notice that equation (4) is linear in $X$. If $X$ is $C^{r+\alpha}$ we can construct a sequence $X_n$ which is analytic in decreasing domains. If $(K_0, \omega_0, \lambda_0)$ is analytic and solves (4) for $X_0$ with a good enough approximation, we can apply Theorem 3.2 and construct a true solution $(K_1, \omega_1, \lambda_1)$ of (4) for $X_0$. This will be an approximate solution of (4) for $X_1$, then applying Theorem 3.2, we can construct an exact solution $(K_2, \omega_2, \lambda_2)$ of the problem for $X_1$, which will be an approximate solution for the problem for $X_2$, etc.

In general, under appropriate inductive assumptions in the domain, we have that

$$||X_n \circ K_n - DK_n A_{\omega_n, \lambda_n}||_{\rho_n} \leq C||X_n - X_{n-1}||_{\rho_n} \leq C2^{-(r+\alpha)n}.$$

Appying Theorem 3.2, we obtain that

$$||K_{n+1} - K_n||_{\rho_{n+1}}, |\omega_{n+1} - \omega_n|, |\lambda_{n+1} - \lambda_n| \leq C2^{-(r+\alpha-1)n}.$$

Hence, we conclude that $K$ is $C^{r-1+\alpha}$.

## 7. Implementation of Algorithm 4.3.

There are several ways to implement Algorithm 4.3. The implementations require choices on the discretization of functions and on the ways to perform the elementary operations. Many practical properties of the algorithm depend on these choices, among them

(a) storage requirements,
(b) speed,
(c) accuracy,
(d) stability,
(e) paralellizability.

Here we will discuss several possible discretizations:

7.1. **Taylor-Fourier series.** This is a very well stablished method in celestial mechanics (see [BG69, Dep70, Sch89] for classical implementations and [LG12, Har08] for more modern implementations).

In this representation, one stores the Fourier-Taylor coefficients of a function

$$f(\theta, s) = \sum_{j \in \mathbb{N}, \, k \in \mathbb{Z}} f_{jk} s^j e^{2\pi i k \theta}. \tag{65}$$

Note that for discretization (65), steps (6) and (7) of Algorithm 4.3 are diagonal and require only $O(N)$ operations. Similarly, the computation of the derivative in step (2), as well as addition and multiplication of functions by numbers, also require only $O(N)$ operations.

The difficult calculation is the computation of $X \circ K$ in step (1). When $X$ can be obtained applying elementary operations (addition, multiplication, trigonometric functions, exponentials, etc.) there is a well defined toolskit that goes under the name of *automatic differentiation.*

It consists of a set of techniques to compute the derivatives of arbitrary order of a function evaluated at a fixed value, accurate to working precision, avoiding in this way the numerical problems inherent in symbolic and numerical differentiation. They are based on writing the function as a sequence of algebraic operations (sum, product,...) and elementary transcendental functions (exp, sin, cos, log, power,...), and then applying systematically the chain rule to these operations (see [Knu97], and also the Web page of the automatic differentiation community http://www.autodiff.org/).

Take for instance the case of the exponential function. Consider given the Taylor expansion of a function $f$, then one can obtain the Taylor expansion of the function $\exp(f)$ using the following relation

$$\partial_s \exp(f) = \exp(f) \partial_s f.$$

Substituting $f$ by its Taylor expansion and equating terms of order $n$ we obtain

$$(n+1)[\exp(f)]_{n+1}(\theta) = \sum_{\ell=0}^{n} [\exp(f)]_{n-\ell}(\theta)(\ell+1) f_{\ell+1}(\theta). \tag{66}$$

We can think of (66) as a recursion that allows us to compute $\exp(f)_{n+1}(\theta)$ provided that we know $\exp(f)_0(\theta), \ldots, \exp(f)_{n-1}(\theta)$. The recursion can be initialized because $\exp(f)_0$ is just the constant $\exp(f_0)$.

Similar algorithms can be obtained for $\sin f$, $\cos f$, $\log(1+f)$, $f^\alpha$, or indeed for any function of $f$ that satisfies a differential equation or some recurrence on the coefficients.

Also, one can apply similar algorithms for Fourier series (think of them as the sums of two polynomials in $e^{2\pi i \theta}$, $e^{-2\pi i \theta}$). Hence, one can

use the previous algorithm to compute

$$\exp\left(\sum_{j\geq 0} f_k(e^{2\pi i\theta})^k\right)\ ,$$

$$\exp\left(\sum_{j\geq 0} f_k(e^{-2\pi i\theta})^k\right)\ ,$$

and then use the addition formula for the exponentials.

### 7.2. **Fourier-real mixed representation.**
A variant of the automatic differentiation which has proved very useful is to keep at the same time both a representation based on the Fourier-Taylor coefficients and a discrete representation in real space

$$f\Big(j/N,\ \exp(2\pi i\ell/N)\Big),$$

for $j = 0, \ldots, N-1$ and $\ell = 0, \ldots, N-1$. Note that one can go from the discrete representation to the Taylor-Fourier series representation just using the Fast Fourier Transform.

Since we use both the Fourier-Taylor representation and the discrete representation we can use the Fourier-Taylor representation in steps (2) – differentiation – and (6),(7) –solving the cohomology equations – of algorithm 4.3 and the discrete representation in step (1) – evaluation of the vector field.

Hence, at the price of doubling the storage space, (which in any case is only $\mathcal{O}(N)$, since we do not use any matrix) we obtain a step which requires only $\mathcal{O}(N\log(N))$ operations, where $N$ is the number of discretization points (or the number of coefficients).

In practice, for most computers, one can find highly optimized implementations of FFT, for example [FJ05] so that the algorithm is $\mathcal{O}(aN\log N + bN)$ with $a \ll b$.

*Remark* 7.1. Note that in order to be able to use the discretization (65) for $K$ we need that the radius of convergence of $K$ is at least 1. As observed in Remark 2.2 this can always be achieved through a proper change of scales.

### 7.3. **Splines.**
When the vector field is not analytic or is given by empirical measurements, a method of choice to discretize the vector field and the function $K$ is to use *"splines"* [dB01]. By now, splines, including multidimensional splines, are well supported in many packages [GDT$^+$11, Eat12].

The discretization in terms of splines takes $\mathcal{O}(N)$ operations to evaluate the vector field, compute derivatives, etc. However, the evaluation of splines makes non-trivial the computation of solutions of cohomology equations (26) and (28). In this case, it is more efficient to use solutions (30) and (31), which require only quadratures. We note that if the spline representation (by polynomials of low order) is known, the quadratures can be computed in closed form and evaluated rather fast (again only $\mathcal{O}(N)$ operations).

Thus, using the splines discretization, a step requires only $O(N)$ operations. Furthermore, given that the operations required for splines are local, they can be easily parallelized, specially in machines with multiple cores. Hence, splines seem to be extremely fast, even for a large number of data points.

7.4. **Numerical norms.** Depending on the discretizations used, the norms that are appropriate may be different.

If we use splines, the error can be measured easily in $C^r$ spaces. Analytic norms are not appropriate for splines since the functions involved are not analytic.

If we use Fourier series, the norms that are easy to compute are the ones that can be expressed in terms of Fourier series. Amongst the most effective norms in Fourier analysis are the weighed $\ell^1$ norms because for them it is easy to compute the norm of operators in terms of matrix elements. For example, $||f||_{w\ell^1} = \sum_k |k|^n |\hat{f}_k| e^{\rho|k|}$ is a norm that has many advantages: it is easy to compute reliably, it is easy to compute for operators and it satisfies the Banach algebra property for $n$ large enough.

For the experts we remark that one could have developed the theoretical results such as Theorem 3.2 in terms of weighted $\ell^1$ norms of the Fourier coefficients, but it turns out that estimates of the composition operator are not so easy. Also the characterization of finite differentiable functions by approximation are only true in the supremum norms considered here.

In practice, one can get useful upper bounds of $||f||_\rho$ by noting that $||f||_\rho \leq ||f||_{w\ell^1}$ or, more generally

$$(67) \qquad\qquad ||f||_\rho \leq \left( ||f^k||_{w\ell^1} \right)^{1/k}.$$

The bounds (67) are very easy to implement and they are very sharp in practice. Indeed, it is a consequence of the theory of Banach algebras [Rud74, Th. 18.9 ] that, for any norm which is a Banach algebra under

multiplication one has

$$||f||_\rho = \lim_{k \to \infty} \left( ||f^k||_{w\ell^1} \right)^{1/k}.$$

## 8. IMPLEMENTATION AND NUMERICAL RESULTS

In this section we discuss some aspects of the numerical implementation using a Fourier-real mixed representation.

**Initial guess:** To apply the Newton method we need to start with an initial approximation for the function $K$ and the parameters $\lambda$ and $\omega$. To do so, we will use a Poincaré section and reduce the problem to finding a zero of the Poincaré map. This will provide $K_0$ (the periodic orbit) and $\omega = 1/T$, where $T$ is the period of the orbit.

A reasonable approximation for $K_1$ and $\lambda$ can be obtained by observing that

$$DX \circ K_0(\theta)K_1(\theta) = \omega D_\theta K_1(\theta) + \lambda K_1(\theta).$$

Hence, we can solve for $U(\theta)$ the variational equation

(68)
$$DX \circ K_0(\theta)U(\theta) = \frac{d}{d\theta}U(\theta)$$
$$U(0) = \mathrm{Id}_2,$$

and we will have that $e^{\lambda/\omega}$ will be an eigenvalue of $U(T)$ and $K_1(0)$ will be the corresponding eigenvector.

Note that once we obtain the result for a point, say $\theta = 0$, it is easy to propagate using that

(69)
$$K_1(\theta) = U(\theta)K_1(0)e^{-\lambda\theta}.$$

Our initial approximation for $K$ will be $K(\theta, \sigma) = K_0(\theta) + K_1(\theta)\sigma$. We store $K_0(\theta)$ and $K_1(\theta)$ for equidistant values of $\theta$; that is, $\theta_i = i/N$ for $i = 0, \ldots, N-1$. Notice that this is equivalent to storing the coefficients of the Fourier series up to degree N by means of the FFT algorithm.

**Newton step:** We will use a Fourier-real mixed representation to implement Algorithm 4.3. See section 7.2 for a detailed description. Since $X$ can be obtained applying elementary operations, we used automatic differentiation methods to perform the Taylor expansions as described in section 7.1. In order to pass from a grid representation to Fourier series and vice versa we use the fast Fourier transform (FFT). In this work we have used the fftw3 library [FJ12, FJ05].

At each step of the Newton method we double the order of the Taylor series, so that after $n$ Newton steps we have computed the Fourier-Taylor series up to order $L = 2^n$.

The Newton method stops when the solution has been computed up to an error (17) of order $10^{-12}$ up to the desired order $L$. The norm we use to estimate the error is the $\ell_1$ norm.

**Local approximation:** Up to this point we assume that we have converged to an approximate solution $K$ of the invariance equation. We need to determine the domain $\Omega_{loc}$ where the solution $K$ is accurate; that is, the function $K$ satisfies the invariance equation up to a certain tolerance $E$ that we established at $10^{-12}$.

Given a fixed tolerance $E$ we compute

$$(70) \quad \Omega_{loc}(E) := \{(\theta, \sigma) \in \mathbb{T} \times \mathbb{R} \mid \|X(K(\theta, \sigma)) - DK(\theta, \sigma) A_{\omega, \lambda}\| < E\},$$

where $\|\cdot\|$ is a norm in $\mathbb{R}^2$. The dependence on $E$ of this domain will be suppressed to simplify notation. We remark that $\Omega_{loc}$ contains the limit cycle $\gamma$. Notice that the higher the order $L$ the larger is $\Omega_{loc}$.

**Globalization of the isochrones:** Since the isochrons are invariant for the time-$T$ map of the flow of the vector field $X$, one can take several points on the isochron provided by the local approximation and iterate these points backwards for the time-$T$ map of the flow $X$. This procedure extends the isochron to a bigger domain $\Omega$. We refer to this procedure as the globalization of isochrons.

We use the procedure described in [GH09], which follows the numerical method proposed in [Sim90]. The main idea is to select a non uniform mesh of points on the isochron so that the globalization procedure provides dense points on the isochron.

The integration method used is a Taylor method (we have used the routines provided by Jorba and Zou; see [18] and http://www.maia.ub.es/angel/soft.html). We used adaptive step size, degree, and a tolerance (absolute and relative) of $10^{-16}$.

8.1. **Numerical Examples.** We present here an application of the method described in the previous section to a classical example in the literature: the Van der Pol oscillator. The equations for the Van der Pol oscillator are:

$$\begin{aligned} \dot{x} &= -y + x - x^3 \\ \dot{y} &= x \end{aligned}$$

The system is analytic, has an unstable fixed point at $(0,0)$ and a stable limit cycle. We have used $N = 2042$ Fourier modes for each $K_n$. The program takes only a few seconds (less than 10s) to compute $K_n$ up to order 60 in a regular laptop.

In Figure 1 we show the limit cycle and 16 isochrones for the Van der Pol oscillator. We show the local approximation up to orders 15, 30 and 60 computed semi-analytically using Fourier-Taylor expansions and the globalized isochron computed using the dynamics given by the vector field.

We already mentioned that the solution is not unique; indeed if $K(\theta, \sigma)$ is a solution of the invariance equation, so is $K(\theta + \theta_0, bs)$ for any $\theta_0 \in [0, 1)$ and $b \in \mathbb{R}$.

We choose $\theta_0$ so that the zero phase for the oscillator corresponds to the maximum value of the $x$-coordinate (in Neuroscience, the peak of the spike).

The choice of $b$ is related to the domain of convergence (70). Hence, if we choose a large $b$, the domain where we can evaluate the series will be small. Although mathematically we can choose any value of $b$, for the numerical stability it will be convenient to choose a value of $b$ such that the coefficients $K_n$ can be kept at order 1, so that one can avoid the round-off errors. However, since the orbits do not approach the limit cycle uniformly along the limit cycle, the functions $K_n$ are not uniform in $\theta$ and one cannot find a global $b$. Hence, for certain values of $\theta$, $K_n$ becomes smaller than the machine precision as $n$ increases. See Figure 1. Given a particular choice of the parameter $b$, increasing the order of the Taylor series has no effect on increasing the length of the local isochron, because the values of $K_n$ are very small. See Figure 1, where we show the computed $K_n$ for some values of $n$.

## References

[Ahl78]  Lars V. Ahlfors. *Complex analysis*. McGraw-Hill Book Co., New York, third edition, 1978. An introduction to the theory of analytic functions of one complex variable, International Series in Pure and Applied Mathematics.

[BG69]  R. Broucke and K. Garthwaite. A programming system for analytical series expansions on a computer. *Celestial Mech.*, 1:271–284, 1969.

[CdlL10]  Renato Calleja and Rafael de la Llave. A numerically accessible criterion for the breakdown of quasi-periodic solutions and its rigorous justification. *Nonlinearity*, 23(9):2029–2058, 2010.

[CFL03a]  Xavier Cabré, Ernest Fontich, and Rafael de la Llave. The parameterization method for invariant manifolds. I. Manifolds associated to non-resonant subspaces. *Indiana Univ. Math. J.*, 52(2):283–328, 2003.
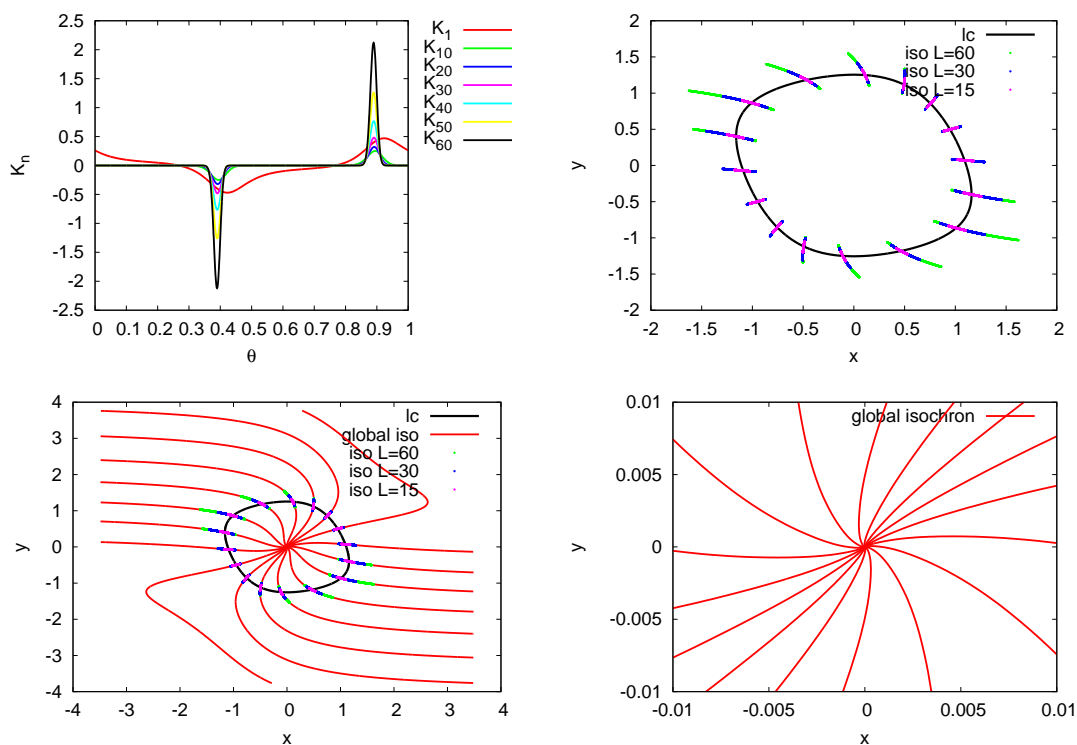
FIGURE 1. The Van der Pol oscillator. (Top left) Functions $K_n$, for $n = 1, 10, 20, 30, 40, 50, 60$. (Top right) Local isochrones obtained from the analytical solution computed up to order 15, 30 and 60. (Bottom Left) Globalization of the local isochrones. (Bottom right) Global isochrones around (0,0).

[CFL03b]  Xavier Cabré, Ernest Fontich, and Rafael de la Llave. The parameterization method for invariant manifolds. II. Regularity with respect to parameters. *Indiana Univ. Math. J.*, 52(2):329–360, 2003.

[CFL05]   Xavier Cabré, Ernest Fontich, and Rafael de la Llave. The parameterization method for invariant manifolds. III. Overview and applications. *J. Differential Equations*, 218(2):444–515, 2005.

[dB01]    Carl de Boor. *A practical guide to splines*, volume 27 of *Applied Mathematical Sciences*. Springer-Verlag, New York, revised edition, 2001.

[Dep70]   A. Deprit. Canonical transformations depending on a small parameter. *Celestial Mech.*, 1:12–30, 1969/1970.

[Eat12]   John W. Eaton. GNU octave 3.6.2. 2012. www.octave.org.

[Erm96]   G. B. Ermentrout. Type I membranes, phase resetting curves, and synchrony. *Neural Comp.*, 8:979–1001, 1996.

[FJ05]    Matteo Frigo and Steven G. Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. Special issue on "Program Generation, Optimization, and Platform Adaptation".

[FJ12]    Matteo Frigo and Steven G. Johnson. FFTW3, 3.3.2. 2012. `fftw.org`.

[GDT+11] M. Galassi, Jim Davis, James Thailer, Brian Gough, Gerard Jungman, Patrick Alken, Michael Booth, and Fabrice Rossi. GSL, gnu scientific library 1.15. 2011. `http://www.gnu.org/software/gsl` .

[GH09]    Antoni Guillamon and Gemma Huguet. A computational and geometric approach to phase resetting curves and surfaces. *SIAM J. Appl. Dyn. Syst.*, 8(3):1005–1042, 2009.

[Guc75]   J. Guckenheimer. Isochrons and phaseless sets. *J. Math. Biol.*, 1(3):259–273, 1974/75.

[Har08]   Alex Haro. Automatic differentiation tools in computational dynamical systems. Manuscript, 2008.

[HL06]    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 (electronic), 2006.

[Izh07]   Eugene M. Izhikevich. *Dynamical systems in neuroscience: the geometry of excitability and bursting*. Computational Neuroscience. MIT Press, Cambridge, MA, 2007.

[JS92]    À. Jorba and C. Simó. On the reducibility of linear differential equations with quasiperiodic coefficients. *J. Differential Equations*, 98(1):111–124, 1992.

[Knu97]   Donald E. Knuth. *The art of computer programming. Vol. 2: Seminumerical algorithms*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont, third revised edition, 1997.

[Kol54]   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. English translation in *Stochastic Behavior in Classical and Quantum Hamiltonian Systems (Volta Memorial Conf., Como, 1977)*, Lecture Notes in Phys., 93, pages 51–56. Springer, Berlin, 1979.

[LG12]    J. Laskar and M. Gastineau. TRIP, a general computer algebra package dedicated to celestial mechanics, 1.2.0. 2012. `http://www.imcce.fr/Equipes/ASD/trip/trip.php`.

[LGJV05]  R. de la Llave, A. González, À. Jorba, and J. Villanueva. KAM theory without action-angle variables. *Nonlinearity*, 18(2):855–895, 2005.

[Lla01]   Rafael de la Llave. A tutorial on KAM theory. In *Smooth ergodic theory and its applications (Seattle, WA, 1999)*, pages 175–292. Amer. Math. Soc., Providence, RI, 2001.

[MC08]    SelvaK. Maran and CarmenC. Canavier. Using phase resetting to predict 1:1 and 2:2 locking in two neuron networks in which firing order is not always preserved. *Journal of Computational Neuroscience*, 24:37–55, 2008.

[Mos66a]  J. Moser. A rapidly convergent iteration method and non-linear differential equations. II. *Ann. Scuola Norm. Sup. Pisa (3)*, 20:499–535, 1966.

[Mos66b]  J. Moser. A rapidly convergent iteration method and non-linear partial differential equations. I. *Ann. Scuola Norm. Sup. Pisa (3)*, 20:265–315, 1966.

[Mos67]   J. Moser. Convergent series expansions for quasi-periodic motions. *Math. Ann.*, 169:136–176, 1967.

[OM10]    Hinke M. Osinga and Jeff Moehlis. Continuation-based computation of global isochrons. *SIAM J. Appl. Dyn. Syst.*, 9(4):1201–1228, 2010. With online multimedia enhancements.

[Poi78]   H. Poincaré. Note sur les propriétés de fonctions définies par les equations diffeérentielles. *Jour. de la Ec. Polyth.*, pages 13–26, 1878.

[Rud74]   Walter Rudin. *Real and complex analysis.* McGraw-Hill Book Co., New York, second edition, 1974. McGraw-Hill Series in Higher Mathematics.

[Sch89]   D.S. Schmidt. Polypak: an algebraic processor for computations in celestial mechanics. In Chudnovsky and pJenks, editors, *Computer Algebra*, pages 111–120. Marcel Dekker, 1989.

[SG09]    W.c Sherwood and J. Guckenheimer. Dissecting the phase response of a model bursting neuron. *arXiv*, 0910.1970, 2009.

[Sim90]   Carles Simó. On the analytical and numerical approximation of invariant manifolds. *Les Méthodes Modernes de la Mecánique Céleste, D. Benest and C. Froeschlé (eds.). Editions Frontières*, pages 285–329, 1990.

[SZ65]    Stanisław Saks and Antoni Zygmund. *Analytic functions.* Second edition, enlarged. Translated by E. J. Scott. Monografie Matematyczne, Tom 28. Państwowe Wydawnietwo Naukowe, Warsaw, 1965. Available from `http://matwbn.icm.edu.pl/kstresc.php?wyd=10&tom=28&jez=en`.

[TF10]    Daisuke Takeshita and Renato Feres. Higher order approximation of isochrons. *Nonlinearity*, 23(6):1303–1323, 2010.

[Win75]   A. T. Winfree. Patterns of phase compromise in biological cycles. *J. Math. Biol.*, 1(1):73–95, 1974/75.

[Zeh75]   E. Zehnder. Generalized implicit function theorems with applications to some small divisor problems/I. *Comm. Pure Appl. Math.*, 28:91–140, 1975.

COURANT INSTITUTE OF MATHEMATICAL SCIENCES, NEW YORK UNIVERSITY, 251 MERCER STREET, NEW YORK, N.Y. 10012-1185
    *E-mail address*: `gh707@nyu.edu`

SCHOOL OF MATHEMATICS, GEORGIA INSTITUTE OF TECHNOLOGY, ATLANTA, GA, 30332, USA
    *E-mail address*: `rafael.delallave@math.gatech.edu`