60 (2015) APPLICATIONS OF MATHEMATICS No. 6, 617–636

ANALYTIC ENCLOSURE OF THE FUNDAMENTAL MATRIX SOLUTION

Roberto Castelli, Amsterdam, Jean-Philippe Lessard, Québec,
Jason D. Mireles James, Boca Raton

(Received June 3, 2015)

Abstract. This work describes a method to rigorously compute the real Floquet normal form decomposition of the fundamental matrix solution of a system of linear ODEs having periodic coefficients. The Floquet normal form is validated in the space of analytic functions. The technique combines analytical estimates and rigorous numerical computations and no rigorous integration is needed. An application to the theory of dynamical system is presented, together with a comparison with the results obtained by computing the enclosure in the C^s category.

Keywords: rigorous numerics; fundamental matrix solution; Floquet theory; analytical category

MSC 2010: 65G99, 34A05, 37B55

1. Introduction

In the theory of linear differential systems, given a homogeneous system of differential equations

$$\dot{y} = A(t)y \tag{1.1}$$

with $A(t) \in M_n(\mathbb{R})$ a τ-periodic matrix valued function, a matrix function $\Phi(t)$ is called a fundamental matrix solution if all columns are linearly independent solutions of (1.1). Here $M_n(\mathbb{K})$ denotes the matrices with entries in the field $\mathbb{K}$. A function $\Phi(t)$ is called a principal fundamental matrix solution if it is a fundamental solution

The third author was partially supported by the National Science Foundation-United States Grant DSM 1318172.

and $\Phi(t_0) = I_n$ for some t_0. Here $I_n \in M_n(\mathbb{R})$ denotes the identity matrix. Among the principal fundamental matrix solutions, we focus on the one that solves

$$\dot{\Phi} = A(t)\Phi, \qquad \Phi(0) = I_n \tag{1.2}$$

and, throughout this paper, we will refer to this as the fundamental matrix solution. Clearly, the fundamental matrix solution alone determines all the solutions of (1.1) in the sense that the orbit $y(t)$ with an initial condition $y(0) = y_0$ is simply given by $y(t) = \Phi(t)y_0$.

Systems of linear differential equations with periodic coefficients are a classical topic of investigation and have applications in a wide range of areas including dynamic stability, elastic systems, Hamiltonian dynamics, celestial mechanics, and engineering systems. See [13] for a survey. Despite the simple formulation, in general it is not possible to write explicitly the solution of system (1.2) in closed form. In addition the numerical integration of the system may produce unreliable results because of the large instabilities introduced by the matrix function $A(t)$.

A significant theoretical tool for studying the fundamental matrix solution is provided by the Floquet theory, which ensures that the function $\Phi(t)$ solving the system (1.2) can be decomposed into the product $\Phi(t) = Q(t)\mathrm{e}^{Rt}$, where $R \in M_n(\mathbb{R})$ and $Q(t) \in M_n(\mathbb{R})$ is a nonsingular, 2τ-periodic matrix valued function. We refer to the latter as the real Floquet normal form decomposition of $\Phi(t)$. Floquet theory identifies the non periodic function $\Phi(t)$ with the couple $(R, Q(t))$, with R a constant matrix and $Q(t)$ a periodic function. The latter can be expanded as a Fourier series, and in this perspective the differential system (1.2) is equivalent to an infinite-dimensional algebraic system where R and the Fourier coefficients of $Q(t)$ are the unknowns. Denoting by $\mathcal{Q} = \{\mathcal{Q}_k\}_{k\in\mathbb{Z}}$ the sequence of Fourier coefficients of $Q(t)$, the problem of solving (1.2) is then rephrased as a zero finding problem $f(x) = 0$ for the unknowns $x = (R, \mathcal{Q})$ in a suitable Banach space $(X, \|\cdot\|)$.

An efficient strategy for obtaining mathematically rigorous enclosures of the solutions of $f(x) = 0$ is given by the radii polynomial approach. This technique is employed in computer-assisted study of many problems in dynamical systems and differential equations, see for instance [8], [7], [4], [10] and the references therein. In short the radii polynomial approach aims at proving the existence of a true solution for f in a certain ball with respect to the norm $\|\cdot\|$ around a numerical approximation. The technique requires rigorous numerical computations, to be performed with the aid of a computer, as well as analytical pen and paper estimates. In this regard, the choice of the norm $\|\cdot\|$ plays an important role: indeed the Banach space determines the regularity of the solution and affects the difficulty of proving sharp estimates.

In [5] the radii polynomial approach is adopted to study system (2.1) in the space of algebraically decaying sequences, i.e. $\|x\| = \max\{\|R\|_\infty, \|\mathcal{Q}\|_s\}$, where

$$\|\mathcal{Q}\|_s \stackrel{\text{def}}{=} \sup_{k\in\mathbb{Z}}\{\|\mathcal{Q}_0\|_\infty, \|\mathcal{Q}_k\|_\infty |k|^s\}, \quad s > 1,$$

hence providing the enclosure of $Q(t)$ in the space of C^s functions. Here, given $M \in M_n(\mathbb{K})$, $\|M\|_\infty$ is the standard ∞ norm, i.e. $\|M\|_\infty = \max_i \sum_j |M_{i,j}|$, where the absolute value sign denotes both the real absolute value (for $\mathbb{K} = \mathbb{R}$) and the complex norm (for $\mathbb{K} = \mathbb{C}$), depending on M. However, even if the solution of the initial value problem (1.2) is a priori known to be analytic, the enclosure in the C^s category does not provide any information about analyticity. The purpose of the present work is to extend the results of [5] and provide the enclosure of the real Floquet normal form decomposition of the fundamental matrix solution in the analytical category. To this end we combine the methodology of [5] with the developments of [9], where the radii polynomial approach is adapted to the study of analytic periodic solutions of differential equations.

The main motivation behind our investigation is our interest in validated computation of analytic parameterizations of stable and unstable manifolds of hyperbolic periodic orbits of vector fields, as presented in [6]. The theoretical foundation of the parameterization method can be found in [1], [2], [3]. The ingredients necessary for computer-assisted validation of the methods of [6] are the analytic representations of both the orbit and the tangent bundle. The latter can be accomplished via analytic representation of the fundamental matrix solution. Analytic representation of the periodic orbit is already provided for instance by [9], and the present work treats the fundamental matrix solution.

We proceed as follows: first we setup the infinite-dimensional algebraic problem we are interested in. Next we briefly review the radii polynomial approach. In Section 3 we focus on the Banach space and provide some preliminary analytical results used later on in Section 4, where the radii polynomial is constructed. Finally in Section 5 we present some computational results and applications to the theory of dynamical systems.

2. Setting of the problem

Aiming at computing the real Floquet normal form decomposition of $\Phi(t)$, we substitute $\Phi(t) = Q(t)\mathrm{e}^{Rt}$ into the equation (1.1). It follows that R and $Q(t)$ solve $\dot{Q}(t) = A(t)Q(t) - Q(t)R$. On the other hand, if $R \in M_n(\mathbb{R})$ and $Q(t) \in M_n(\mathbb{R})$,

a 2τ-periodic matrix function, solve

$$\begin{cases} \dot{Q}(t) = A(t)Q(t) - Q(t)R, \\ Q(0) = I_n, \end{cases} \tag{2.1}$$

then $\Phi(t) = Q(t)\mathrm{e}^{Rt}$ is the fundamental matrix solution. By assumption, the matrix function $A(t)$ is a given τ-periodic function. Hence it is 2τ-periodic and it admits a Fourier series expansion of the form

$$A(t) = \sum_{k\in\mathbb{Z}} \mathcal{A}_k \mathrm{e}^{\mathrm{i}k2\pi t/(2\tau)}, \quad \mathcal{A}_k \in M_n(\mathbb{C}). \tag{2.2}$$

Let

$$Q(t) = \sum_{k\in\mathbb{Z}} \mathcal{Q}_k \mathrm{e}^{\mathrm{i}k2\pi t/(2\tau)}, \quad \mathcal{Q}_k \in M_n(\mathbb{C}) \tag{2.3}$$

be the Fourier decomposition of the 2τ-periodic unknown function $Q(t)$ and denote by $\mathcal{Q} \stackrel{\mathrm{def}}{=} \{\mathcal{Q}_k\}_{k\in\mathbb{Z}}$ the sequence of the Fourier coefficients of $Q(t)$. After projecting into the Fourier space the ODE (2.1) is equivalent to the infinite-dimensional algebraic system

$$\begin{gathered} f(R, \mathcal{Q}) = 0 \\ f = (\dots, f_{-k}, \dots, f_{-1}, f_\star, f_0, f_1, \dots, f_k, \dots) \end{gathered} \tag{2.4}$$

defined by

$$\begin{aligned} f_\star &\stackrel{\mathrm{def}}{=} \sum_{k\in\mathbb{Z}} \mathcal{Q}_k - I_n, \\ f_k &\stackrel{\mathrm{def}}{=} \mathrm{i}k\frac{2\pi}{2\tau}\mathcal{Q}_k + \mathcal{Q}_k R - (\mathcal{A} * \mathcal{Q})_k, \quad k \in \mathbb{Z}, \end{aligned} \tag{2.5}$$

where $(\mathcal{A} * \mathcal{Q})_k$ denotes the convolution product $(\mathcal{A} * \mathcal{Q})_k \stackrel{\mathrm{def}}{=} \sum\limits_{k_1+k_2=k} \mathcal{A}_{k_1}\mathcal{Q}_{k_2}$. Hence, solving system (2.1) is equivalent to looking for zeros of the algebraic system f in the unknowns $(R, \{\mathcal{Q}_k\}_{k\in\mathbb{Z}})$. We introduce the Banach space

$$\begin{gathered} X \stackrel{\mathrm{def}}{=} \{x = (R, \mathcal{Q}) \colon \|x\|_X \stackrel{\mathrm{def}}{=} \max\{\|R\|_\infty, \|\mathcal{Q}\|_{1,\nu}\} < \infty\} \\ \text{where } \|\mathcal{Q}\|_{1,\nu} \stackrel{\mathrm{def}}{=} \sum_{k\in\mathbb{Z}} \|\mathcal{Q}_k\|_\infty \nu^{|k|}, \ \nu \geqslant 1. \end{gathered}$$

The quantity $\|\mathcal{Q}\|_{1,\nu}$ is a weighted l^1 norm on the space of sequences of complex valued matrices. Denote

$$l^1_\nu(M_n(\mathbb{C})) \stackrel{\text{def}}{=} \{\mathcal{Q} = \{\mathcal{Q}_k\}_{k\in\mathbb{Z}}\colon\ \mathcal{Q}_k \in M_n(\mathbb{C}),\ \|\mathcal{Q}\|_{1,\nu} < \infty\}.$$

In the following we simply denote $l^1_\nu(M_n(\mathbb{C}))$ as l^1_ν.

Note that a sequence in l^1_ν has an exponential decay rate. That makes l^1_ν a suitable norm in the analytic function space. Indeed, if $(R, \mathcal{Q}) \in X$ is a solution of $f(R, \mathcal{Q}) = 0$ so that $\|(R, \mathcal{Q})\|_X < \infty$, the corresponding function $Q(t)$ is analytic in a complex strip around the real line (e.g. see [11]).

For a finite-dimensional projection parameter m and $x = (R, \{\mathcal{Q}_k\}_{k\in\mathbb{Z}})$, we define

$$x^{(m)} = (R, \{\mathcal{Q}_k\}_{|k|<m}) \quad \text{and} \quad f^{(m)} = (f_\star, \{f_k\}_{|k|<m}).$$

Suppose that a numerical solution $\bar{x} = (\overline{R}, \{\overline{\mathcal{Q}}_k\}_{|k|<m})$ has been computed, that is $f^{(m)}(\bar{x}) \approx 0$. Let $Df^{(m)}(\bar{x})$ be the derivative of $f^{(m)}$ with respect to $x^{(m)}$ at $\bar{x}$ and

$$\Lambda_k \stackrel{\text{def}}{=} \frac{\partial f_k}{\partial \mathcal{Q}_k}(\bar{x}). \tag{2.6}$$

Consider $J^{(m)}$, an approximate inverse (computed numerically) of $Df^{(m)}(\bar{x})$ and let J be the operator

$$(Jx)_k \stackrel{\text{def}}{=} \begin{cases} (J^{(m)}x^{(m)})_k, & k = \star,\ |k| < m, \\ \Lambda_k^{-1}x_k, & |k| \geqslant m. \end{cases} \tag{2.7}$$

By construction, $J\colon X \to X$ acts as an approximation of $Df^{-1}(\bar{x})$. Assume J is injective and define the Newton-like operator

$$T(x) = x - Jf(x) \tag{2.8}$$

so that the fixed points for T correspond to the solutions of $f = 0$. The core of the radii polynomial approach consists in defining a bound Y and a polynomial bound $Z(r)$ satisfying

$$\|T(\bar{x}) - \bar{x}\|_X \leqslant Y \quad \text{and} \quad \sup_{b,c\in B(r)} \|DT(\bar{x} + b)c\|_X \leqslant Z(r) \tag{2.9}$$

and the radii polynomial

$$p(r) \stackrel{\text{def}}{=} Y + Z(r) - r.$$

Finally, as motivated in the following Lemma, we check if $p(r) < 0$ for some r.

Lemma 2.1. *Let Y, $Z(r)$ be chosen such that the inequalities* (2.9) *are satisfied and let $p(r)$ be defined as above. If there exists $r_0 > 0$ such that $p(r_0) > 0$ then there exists a unique $\tilde{x} \in \overline{B_{r_0}(\bar{x})} \stackrel{\text{def}}{=} \{x \in X\colon \|x - \bar{x}\|_X \leqslant r_0\}$ such that $\tilde{x} = T(\tilde{x})$, or equivalently such that $f(\tilde{x}) = 0$.*

P r o o f. See for instance [5]. □

R e m a r k 2.1. To ensure that the function $Q(t)$ is a real one needs that the coefficients $\mathcal{Q}_k$ satisfy the conjugacy symmetry $\mathcal{Q}_{-k} = \operatorname{conj}(\mathcal{Q}_k)$. Here $\operatorname{conj}(M)$ is the matrix obtained by taking the component-wise complex conjugate entries of $M \in M_n(\mathbb{C})$. We do not impose such symmetry to the space X. However, if the operator J preserves the symmetry then the fixed point of T is symmetric and the solution $Q(t)$ real.

2.1. Assumptions on the matrix function $A(t)$. Let us now make more explicit assumptions on the matrix function $A(t)$. We assume that the Fourier coefficients of $\mathcal{A}_k$ in the expansion (2.2) are given within certain bounds and have the following properties:

(1) $\mathcal{A}_{-k} = \operatorname{conj}(\mathcal{A}_k)$.
(2) There exist M_A, ν, $r_A > 0$ and $\mathbb{A} = \{\mathbb{A}_k\}_{k\in\mathbb{Z}}$ with $\mathbb{A}_k \in M_n(\mathbb{C})$ such that $\mathbb{A}_k = 0$ for $|k| \geqslant M_A$ and that $\|\mathcal{A} - \mathbb{A}\|_{1,\nu} \leqslant r_A$. In other words,

$$\mathcal{E} \stackrel{\text{def}}{=} \mathcal{A} - \mathbb{A} \text{ satisfies } \|\mathcal{E}\|_{1,\nu} \leqslant r_A. \tag{2.10}$$

The sequence $\mathcal{E} = \{\mathcal{E}_k\}$ is the error bound for $\mathcal{A}$ and it refers to the fact that the function $A(t)$ might not be known exactly. Indeed, in the applications, $A(t)$ may result from measurements, or may be subjected to random noise, or may depend on some data previously computed and given within some bounds only.

3. Analityc preliminaries

To begin with, we remark that l^1_ν is a Banach algebra under the discrete convolution product. In particular, the following result holds.

Lemma 3.1. *Let $V, U \in l^1_\nu$ and let $V * U$ be the discrete convolution product, $(V * U)_k = \sum\limits_{k_1+k_2=k} V_{k_1} U_{k_2}$. Then*

$$\|V * U\|_{1,\nu} \leqslant \|V\|_{1,\nu} \|U\|_{1,\nu}.$$

P r o o f.

$$\begin{aligned}\|V*U\|_{1,\nu} &= \sum_{k\in\mathbb{Z}}\Big\|\sum_{k_1+k_2=k} V_{k_1}U_{k_2}\Big\|_\infty \nu^{|k|} = \sum_{k\in\mathbb{Z}}\Big\|\sum_{l\in\mathbb{Z}} V_l U_{k-l}\Big\|_\infty \nu^{|k|}\\ &\leqslant \sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} \|V_l\|_\infty \|U_{k-l}\|_\infty \nu^{|k|} \leqslant \sum_{l\in\mathbb{Z}} \|V_l\|_\infty \nu^{|l|} \sum_{k\in\mathbb{Z}} \|U_{k-l}\|_\infty \nu^{|k-l|}\\ &\leqslant \|V\|_{1,\nu}\|U\|_{1,\nu}.\end{aligned}$$

□

For a sequence $B=\{B_k\}_{k\in\mathbb{Z}}$ with $B_k\in M_n(\mathbb{C})$, denote the weighted l^∞ norm on the space of sequences of complex valued matrices by

$$\|B\|_{\infty,\nu^{-1}} \stackrel{\text{def}}{=} \sup_{k\in\mathbb{Z}}\Big\{\frac{\|B_k\|_\infty}{\nu^{|k|}}\Big\}.$$

Let

$$l^\infty_{\nu^{-1}}(M_n(\mathbb{C})) \stackrel{\text{def}}{=} \{B=\{B_k\}_{k\in\mathbb{Z}}\colon\ B_k\in M_n(\mathbb{C}),\ \|B\|_{\infty,\nu^{-1}}<\infty\}.$$

In the following we simply denote $l^\infty_{\nu^{-1}}(M_n(\mathbb{C}))$ as $l^\infty_{\nu^{-1}}$.

Lemma 3.2. *Let $V\in l^1_\nu$ and $B\in l^\infty_{\nu^{-1}}$. Then*

$$\Big\|\sum_{k\in\mathbb{Z}} B_kV_k\Big\|_\infty \leqslant \|B\|_{\infty,\nu^{-1}}\|V\|_{1,\nu}.$$

P r o o f.

$$\begin{aligned}\Big\|\sum_{k\in\mathbb{Z}} B_kV_k\Big\|_\infty &\leqslant \sum_{k\in\mathbb{Z}}\|B_kV_k\|_\infty \leqslant \sum_{k\in\mathbb{Z}}\frac{\|B_k\|_\infty}{\nu^{|k|}}\|V_k\|_\infty\nu^{|k|}\\ &\leqslant \sup_{k\in\mathbb{Z}}\frac{\|B_k\|_\infty}{\nu^{|k|}}\Big(\sum_{k\in\mathbb{Z}}\|V_k\|_\infty\nu^{|k|}\Big) \leqslant \|B\|_{\infty,\nu^{-1}}\|V\|_{1,\nu}.\end{aligned}$$

□

For any $B\in l^\infty_{\nu^{-1}}$ the linear operator $\mathcal{L}_B\colon\ l^1_\nu\to(M_n(\mathbb{C}),\|\cdot\|_\infty)$ defined as

$$\mathcal{L}_B(V)=\sum_{k\in\mathbb{Z}} B_kV_k \tag{3.1}$$

is well-defined, and Lemma 3.2 states that

$$|\!|\!|\mathcal{L}_B|\!|\!| \stackrel{\text{def}}{=} \sup_{\|V\|_{1,\nu}\leqslant 1}\|\mathcal{L}_B(V)\|_\infty \leqslant \|B\|_{\infty,\nu^{-1}}.$$

We refer to $l^\infty_{\nu^{-1}}$ as the dual space of l^1_ν.

Corollary 3.1. *Let $\nu \geqslant 1$ and $V \in l^1_\nu$. Then for $m \in \mathbb{N}$,*

$$\left\| \sum_{|k| \geqslant m} V_k \right\|_\infty \leqslant \frac{1}{\nu^m} \|V\|_{1,\nu}.$$

P r o o f. $\sum\limits_{|k|\geqslant m} V_k = \mathcal{L}_B(V)$ where $B_k = 0$ for $|k| < m$ and $B_k = I$ for $|k| \geqslant m$. Then the result follows from Lemma 3.2. □

Definition 3.1. Given $M\colon\ l^1_\nu \to l^1_\nu$ a linear operator, define the operator norm as

$$|\!|\!| M |\!|\!| = \sup_{\|V\|_{1,\nu} \leqslant 1} \|M(V)\|_{1,\nu}.$$

With a linear operator $M\colon\ l^1_\nu \to l^1_\nu$ we associated an infinite-dimensional matrix (still denoted by M) $M = \{M^k_l\}_{k,l\in\mathbb{Z}}$ with $M^k_l \in M_n(\mathbb{C})$, such that

$$(M(V))_k = \sum_{l\in\mathbb{Z}} M^k_l(V_l).$$

Lemma 3.3. *Let $M = \{M^k_l\}_{k,l\in\mathbb{Z}}$ be the matrix representation of an operator $M\colon\ l^1_\nu \to l^1_\nu$. Suppose that any linear operator M^k_l acts as a matrix multiplication. Then*

$$|\!|\!| M |\!|\!| = \sup_{l\in\mathbb{Z}} \frac{1}{\nu^{|l|}} \sum_{k\in\mathbb{Z}} \|M^k_l\|_\infty \nu^{|k|}$$

and for any $V \in l^1_\nu$,

$$\|M(V)\|_{1,\nu} \leqslant |\!|\!| M |\!|\!| \|V\|_{1,\nu}.$$

The following corollary is a direct consequence of the previous lemma and concerns operators whose matrix representation acts as a diagonal multiplication out of a finite-dimensional core.

For $N > 0$ denote $I_N \stackrel{\text{def}}{=} \mathbb{Z}^2 \setminus \{(k,l)\colon\ |k| \leqslant N,\ |l| \leqslant N\}$.

Lemma 3.4. *Let $V = \{V_k\}_{k\in\mathbb{Z}} \in l^1_\nu$ and let $M\colon\ l^1_\nu \to l^1_\nu$ be an operator whose matrix representation $M = \{M^k_l\}$ is such that*

- ▷ $M^k_l = 0\ \forall (k,l) \in I_N \cap \{k \neq l\}$,
- ▷ $M^k_k = \Gamma_k \in M_n(\mathbb{C})\ \forall\, k > N$.

Assume that $\|\Gamma_k\|_\infty \leqslant \gamma_N$ for all $|k| > N$ for some $\gamma_N > 0$. Define

$$K_0 \stackrel{\text{def}}{=} \sum_{j=-N}^{N} \|M^j_0\|_\infty \nu^{|j|}, \quad K_k = \frac{1}{\nu^k} \left(\sum_{j=-N}^{N} \|M^j_k\|_\infty \nu^{|j|} \right)$$

and

$$K = \max\{K_0, \max_{k=-N,\dots,N} K_k\}.$$

Then

$$\|\!|M|\!\| \leqslant \max\{K, \gamma_N\}.$$

4. Construction of the radii polynomial

In the following we continuously employ a correspondence between the space X and the space of bi-infinite sequences of matrices. In practice, the components of an element $(R, \mathcal{Q}) \in X$ are rearranged in the vector form, i.e.

$$V = [\dots, V_{-k} \dots, V_{-1}, V_\star, V_0, V_1, \dots, V_k, \dots],$$

where $V_\star = R$ and $V_k = \mathcal{Q}_k$. Similarly, a linear operator $M\colon X \to X$ can be seen as the action of a matrix of operators (still denoted by M) against a vector V. The same notation is used to label the rows and the columns of the matrix M, respectively, by superscript and subscript:

$$M = \begin{bmatrix} \vdots \\ M^{-k} \\ \vdots \\ M^{-1} \\ M^\star \\ M^0 \\ \vdots \\ M^k \\ \vdots \end{bmatrix}, \quad M = [\dots, M_{-k}, \dots, M_\star, M_0, \dots, M_k, \dots], \quad M_b^a = (M^a)_b.$$

Let us now study the differential of $f(R, \mathcal{Q})$, necessary for the definition of the fixed-point operator T defined in (2.8). The derivative of f at $(R, \mathcal{Q})$ in the direction of $(\alpha, \beta) \in X$, where $\beta = \{\beta_k\}_{k\in\mathbb{Z}} \in l^1_\nu$ is

$$Df_\star(R, \mathcal{Q})(\alpha, \beta) = \sum_{k\in\mathbb{Z}} \beta_k,$$

$$Df_k(R, \mathcal{Q})(\alpha, \beta) = \mathrm{i}k\frac{2\pi}{2\tau}\beta_k + \beta_k R + \mathcal{Q}_k\alpha - (\mathcal{A} * \beta)_k \quad \forall\, k \in \mathbb{Z}.$$

In matrix notation, the derivative $Df(R, \mathcal{Q})(\alpha, \beta)$ results by applying to the vector

$$V = [\dots, \beta_{-k}, \dots, \beta_{-1}, \alpha, \beta_0, \beta_1, \dots, \beta_k, \dots]^{\mathrm{T}}$$

the Jacobian operator Jf given as follows:

$$Jf_{\star,\star} = 0, \quad Jf_{\star,k} = I, \quad Jf_{k,\star} = \mathcal{Q}_k \quad \forall\, k \in \mathbb{Z}, \tag{4.1}$$
$$Jf_{k,j}\colon\ M \to \Big(\mathrm{i}k\frac{2\pi}{2\tau}M + MR\Big)\delta_{k,j} - \mathcal{A}_{k-j}M \quad \forall\, k, j \in \mathbb{Z}.$$

R e m a r k 4.1. Since the matrix multiplication is not commutative, we cannot simply write the Jacobian element $Jf_{k,j}$ as a matrix.

R e m a r k 4.2. From the computational point of view, we represent the $n \times n$ matrices as n^2 vector in such a way that $A = [a_{i,j}]$ is represented by the vector $V_a = [a_{1,1}, \dots, a_{n,1}, a_{1,2}, \dots a_{n,2}, \dots]^{\mathrm{T}}$ (by columns). Given a matrix B, the multiplication $B \cdot A$ is represented by the multiplication $\widehat{B} \cdot V_a$ where $\widehat{B}$ is a block diagonal concatenation of n copies of B, while the right multiplication $A \cdot B$ is given by a multiplication of V_a by the Kronecker product of the transpose of B with the identity I_n.

Once a representation of the $(n \times n)$-matrix into an n^2 vector is chosen, the operators $Jf_{k,j}$ can be represented as a multiplication of an $n^2 \times n^2$ matrix. The operator Λ_k and its inverse Λ_k^{-1} are then represented by a matrix as well. When we write $\|\Lambda_k\|_\infty$ and $\|\Lambda_k^{-1}\|_\infty$ we refer to the inf-norm of the matrix representing the operator.

Lemma 4.1. *Recall* (2.6). *For m sufficiently large and depending only on the data of the problem, there exists $\lambda_m > 0$ such that $\|\Lambda_k^{-1}\|_\infty \leqslant 1/\lambda_m$ for any $|k| \geqslant m$.*

P r o o f. This follows from the fact that for k large enough any matrix Λ_k is diagonal dominant and the absolute values of the elements on the diagonal increase with k. More precisely, in [5] it is proved that there exist $M > 0$ and $C_\Lambda > 0$ such that $\|\Lambda_k^{-1}\|_\infty \leqslant C_\Lambda/k$ for any $|k| \geqslant M$. Letting $\lambda_m = m/C_\Lambda$, we can conclude that $\|\Lambda_k^{-1}\|_\infty \leqslant C_\Lambda/k \leqslant 1/\lambda_m$ for all $k \geqslant m \geqslant M$. □

Assume that the finite dimensional parameter m is chosen according to the previous lemma, then the operator J introduced in (2.7) is injective. Indeed, by construction, the finite dimensional matrix $J^{(m)}$ is invertible and the injectivity of the infinite dimensional tail is ensured by the lemma.

We are now concerned with the definition of the bounds Y and $Z(r)$. These bounds involve the Fourier coefficients $\mathcal{A}_k$ of the function $A(t)$ that are only known within the bounds (2.10). Whenever possible, we separate the contribution of the centre $\mathbb{A}$ from the one of the error $\mathcal{E}$ to obtain the sharpest estimate for Y and $Z(r)$.

4.1. Bound Y. We have

$$T(\bar{x}) - \bar{x} = Jf(\bar{x}),$$

hence define

$$Y = \max\{\|(Jf(\bar{x}))_\star\|_\infty, \|Jf(\bar{x})\|_\nu\}.$$

By inserting $\mathcal{A} = \mathbb{A} + \mathcal{E}$ in (2.5), we write $f(\bar{x}) = \overline{f}(\bar{x}) + \mathcal{E}_f(\bar{x})$ where

$$\overline{f}_\star(\bar{x}) \stackrel{\text{def}}{=} \sum_{|k|<m} \overline{\mathcal{Q}}_k - I_n, \quad \overline{f}_k(\bar{x}) \stackrel{\text{def}}{=} \mathrm{i}k\frac{2\pi}{2\tau}\overline{\mathcal{Q}}_k + \overline{\mathcal{Q}}_k\overline{R} - (\mathbb{A} * \overline{\mathcal{Q}})_k, \quad k \in \mathbb{Z}, \tag{4.2}$$

and

$$(\mathcal{E}_f(\bar{x}))_\star = 0, \quad (\mathcal{E}_f(\bar{x}))_k = -(\mathcal{E} * \overline{\mathcal{Q}})_k, \quad k \in \mathbb{Z}. \tag{4.3}$$

Therefore,

$$(Jf(\bar{x}))_\star = J^\star \cdot \overline{f}(\bar{x}) + J^\star \cdot \mathcal{E}_f(\bar{x}), \quad (Jf(\bar{x}))_k = [J\overline{f}(\bar{x})]_k + [J\mathcal{E}_f(\bar{x})]_k, \quad k \in \mathbb{Z},$$

and

$$\begin{aligned}\|(Jf(\bar{x}))_\star\|_\infty &\leqslant \|J^\star \cdot \overline{f}(\bar{x})\|_\infty + \|J^\star \cdot \mathcal{E}_f(\bar{x})\|_\infty,\\ \|(Jf(\bar{x}))\|_{1,\nu} &\leqslant \|(J\overline{f}(\bar{x}))\|_{1,\nu} + \|(J\mathcal{E}_f(\bar{x}))\|_{1,\nu}.\end{aligned}$$

At this point we recall the assumptions discussed in Section 2.1. In particular we recall that $\mathbb{A}$ is finite-dimensional, i.e. $\mathbb{A}_k = 0$ for $|k| \geqslant M_A$, and that $\|\mathcal{E}\|_{1,\nu} \leqslant r_A$. Since $\overline{Q}$ is such that $\overline{\mathcal{Q}}_k = 0$ for $|k| \geqslant m$, the coefficients $\overline{f}_k$ vanish for $|k| \geqslant M_A + m - 1$. On the other hand, $(\mathcal{E}_f(\bar{x}))_k$ may be different from zero for any k and Lemma 3.1 implies that $\|\mathcal{E}_f(\bar{x})\|_{1,\nu} \leqslant r_A\|\overline{Q}\|_{1,\nu}$. We treat the operations on $\mathcal{E}_f(\bar{x})$ as linear operators and we use the space duality to bound the norms.

From Lemma 3.2 we have

$$\begin{aligned}\|J^\star \cdot \overline{f}(\bar{x})\|_\infty + \|J^\star \cdot \mathcal{E}_f(\bar{x})\|_\infty \leqslant &\left\| J^\star_\star \overline{f}_\star(\bar{x}) + \sum_{|k|<m} J^\star_k \overline{f}_k(\bar{x}) \right\|_\infty\\ &+ \|J^\star\|_{\infty,\nu^{-1}} r_A \|\overline{Q}\|_{1,\nu} =: Y^\star.\end{aligned}$$

Similarly, a bound for $\|(Jf(\bar{x}))\|_{1,\nu}$ is

$$\|(Jf(\bar{x}))\|_{1,\nu} \leqslant \sum_{|k|<M_A+m-1} \|(J\overline{f}(\bar{x}))_k\|_\infty \nu^{|k|} + |\!|\!| J |\!|\!| \, r_A \|\overline{\mathcal{Q}}\|_{1,\nu} =: Y^\nu,$$

where $|\!|\!| J |\!|\!|$ is estimated as shown in Lemma 3.4. Finally, we define $Y = \max\{Y^\star, Y^\nu\}$.

4.2. Bound Z. We are now concerned with the construction of the bound $Z(r)$. Treating r as a variable, we aim at defining a polynomial bound $Z(r)$ such that the second of the inequalities (2.9) is satisfied for any positive r.

Let

$$(J^\dagger x)_k \stackrel{\text{def}}{=} \begin{cases} (Df^{(m)}(\bar{x}) \cdot x^{(m)})_k, & k = \star,\ |k| < m, \\ \Lambda_k x_k, & |k| \geqslant m, \end{cases} \tag{4.4}$$

and consider the splitting

$$\|DT(\bar{x}+b)c\|_X \leqslant \|(I - JJ^\dagger)c\|_X + \|J[(Df(\bar{x}+b) - J^\dagger)c]\|_X.$$

The bound $Z(r)$ is constructed as a polynomial in the variable r as

$$Z(r) = Z^{(0)}r + Z^{(1)}r + Z^{(2)}r^2,$$

where $Z^{(0)}, Z^{(1)}$ and $Z^{(2)}$ are defined so as to satisfy

$$\sup_{c \in B(r)} \|(I - JJ^\dagger)c\|_X \leqslant Z^{(0)}r, \quad \sup_{b,c \in B(r)} \|J[(Df(\bar{x}+b) - J^\dagger)c]\|_X \leqslant Z^{(1)}r + Z^{(2)}r^2.$$

In order to compute the above bounds $Z^{(i)}$, we first factor out r writing $b = ru$, $c = rv$ for $u = [u_\star, \{u_k\}_{k\in\mathbb{Z}}]$ and $v = [v_\star, \{v_k\}_{k\in\mathbb{Z}}]$, both in $\overline{B_1(0)} = \{x \in X\colon \|x\|_X \leqslant 1\}$. That means $\|u_\star\|_\infty \leqslant 1, \|u\|_{1,\nu} \leqslant 1$ and the same for v. From [5] we have that

$$[(Df(\bar{x}+ru) - J^\dagger)rv]_k = \sum_{i=1,2} c_{k,i} r^i, \tag{4.5}$$

where

$$c_{\star,1} = \sum_{|k| \geqslant m} v_k, \quad c_{\star,2} = 0, \tag{4.6}$$

$$c_{k,1} = -\sum_{\substack{k_1+k_2=k \\ |k_2| \geqslant m}} \mathcal{A}_{k_1} v_{k_2}, \quad c_{k,2} = u_k v_\star + v_k u_\star, \quad |k| < m, \tag{4.7}$$

$$c_{k,1} = -\sum_{\substack{k_1+k_2=k \\ k_2 \neq k}} \mathcal{A}_{k_1} v_{k_2}, \quad c_{k,2} = u_k v_\star + v_k u_\star, \quad |k| \geqslant m.$$

4.2.1. Compute $Z^{(0)}$. By definition, $Z^{(0)}$ is constructed so that

$$Z^{(0)} \geqslant \sup_{\|v\|_X \leqslant 1} \|(I - JJ^\dagger)v\|_X.$$

Note that in the finite part $J^\dagger$ acts as the multiplication by $Df^{(m)}(\bar{x})$. The latter, since depends on the coefficients $\mathcal{A}_k$, is known only within bounds. Hence we decompose $Df^{(m)}(\bar{x})$ in the sum $Df^{(m)}(\bar{x}) = \overline{Df}^{(m)}(\bar{x}) + \mathcal{E}_{Df}$ where the dependence of $\overline{Df}^{(m)}$ on $\mathcal{A}$ is only through $\mathbb{A}$ and the operator $\mathcal{E}_{Df}\colon\ \beta^{(m)} \to -(\mathcal{E} * \beta^{(m)})^{(m)}$.

Accordingly, $(I - JJ^\dagger)v = (I^{(m)} - J^{(m)}\overline{Df}^{(m)} - J^{(m)}\mathcal{E}_{Df})v^{(m)}$. Denote $B = I^{(m)} - J^{(m)}\overline{Df}^{(m)}$. We have

$$\|(I - JJ^\dagger)v\|_X \leqslant \|Bv^{(m)}\|_X + \|J^{(m)}\mathcal{E}_{Df}v^{(m)}\|_X. \tag{4.8}$$

By definition $\|Bv^{(m)}\|_X = \max\{\|(Bv^{(m)})_\star\|_\infty, \|Bv^{(m)}\|_{1,\nu}\}$ and Lemma 3.2 yields

$$\begin{aligned}\|(Bv^{(m)})_\star\|_\infty &\leqslant \|B^\star_\star v_\star\|_\infty + \Big\|\sum_{|j|<m} B^\star_j v_j\Big\|_\infty \\ &\leqslant \|B^\star_\star\|_\infty\|v_\star\|_\infty + \|B^\star\|_{\infty,\nu^{-1}}\|v^{(m)}\|_{1,\nu}.\end{aligned} \tag{4.9}$$

Let us now estimate

$$\|Bv^{(m)}\|_{1,\nu} = \sum_{|k|<m} \|(Bv^{(m)})_k\|_\infty \nu^{|k|}.$$

Denote by $\widetilde{B}$ the submatrix of B obtained by deleting the row $B^\star$ and the column $B_\star$ and let $\tilde{v} = \{v_k\}_{|k|<m}$, that is $v^{(m)}$ without $v_\star$. From Lemma 3.3

$$\begin{aligned}\|Bv^{(m)}\|_{1,\nu} &\leqslant \sum_{|k|<m} \|B^k_\star v_\star\|_\infty \nu^{|k|} + \|\widetilde{B}\tilde{v}\|_{1,\nu} \\ &\leqslant \sum_{|k|<m} \|B^k_\star\|_\infty\|v_\star\|_\infty\nu^{|k|} + |\!|\!|\widetilde{B}|\!|\!|\,\|\tilde{v}\|_{1,\nu} = \|B_\star\|_{1,\nu}\|v_\star\|_\infty + |\!|\!|B|\!|\!|\,\|v^{(m)}\|_{1,\nu}.\end{aligned}$$

Altogether,

$$\begin{aligned}\|Bv^{(m)}\|_X \leqslant \max\{&\|B^\star_\star\|_\infty\|v_\star\|_\infty + \|B^\star\|_{\infty,\nu^{-1}}\|v^{(m)}\|_{1,\nu}, \\ &\|B_\star\|_{1,\nu}\|v_\star\|_\infty + |\!|\!|B|\!|\!|\,\|v^{(m)}\|_{1,\nu}\}.\end{aligned}$$

In order to bound the last term in (4.8), we note that $(\mathcal{E}_{Df}v^{(m)})_\star = 0$ and $(\mathcal{E}_{Df}v^{(m)})_k = -(\mathcal{E} * v^{(m)})_k$. In particular, Lemma 3.1 infers that $\|\mathcal{E}_{Df}v^{(m)}\|_{1,\nu} \leqslant r_A\|v^{(m)}\|_{1,\nu}$. Hence, Lemma 3.2 and Lemma 3.3 imply

$$\begin{aligned}\|(J^{(m)}\mathcal{E}_{Df}v^{(m)})_\star\|_\infty &\leqslant \|J^\star\|_{\infty,\nu^{-1}} r_A\|v^{(m)}\|_{1,\nu}, \\ \|(J^{(m)}\mathcal{E}_{Df}v^{(m)})\|_{1,\nu} &\leqslant |\!|\!|J^{(m)}|\!|\!|\, r_A\|v^{(m)}\|_{1,\nu}.\end{aligned}$$

Collecting all the terms and taking the sup over $\|v\|_X \leqslant 1$, we define

$$Z^0 \stackrel{\text{def}}{=} \max\{\|B^\star_\star\|_\infty + \|B^\star\|_{\infty,\nu^{-1}}, \|B_\star\|_{1,\nu} + |\!|\!|B|\!|\!|\} + r_A \max\{\|J^\star\|_{\infty,\nu^{-1}}, |\!|\!|J^{(m)}|\!|\!|\}.$$

4.2.2. Compute $Z^{(1)}$. According to (4.5), $Z^{(1)}$ is defined so that

$$Z^{(1)} \geqslant \sup_{\|v\|_X \leqslant 1} \|J[(Df(\bar{x}) - J^\dagger)v]\|_X = \sup_{\|v\|_X \leqslant 1} (\max\{\|[JC]_\star\|_\infty, \|JC\|_{1,\nu}\}),$$

where $C = (c_{\star,1}, \{c_{k,1}\})$, defined in (4.7), depends on v.

We provide first a bound for $[JC]_\star$. First,

$$[JC]_\star = J^\star_\star \sum_{|j|\geqslant m} v_j - \sum_{|k|<m} J^\star_k (\mathcal{A} * v^{(I)})_k,$$

where $v^{(I)}$ stands for the sequence v with $v_k = 0$ for $|k| < m$. We insert the splitting $\mathcal{A} = \mathbb{A} + \mathcal{E}$

$$[JC]_\star = J^\star_\star \sum_{|j|\geqslant m} v_j - \sum_{|k|<m} J^\star_k (\mathbb{A} * v^{(I)})_k - \sum_{|k|<m} J^\star_k (\mathcal{E} * v^{(I)})_k$$

and compute

$$\begin{aligned}\|[JC]_\star\|_\infty &\leqslant \left\| J^\star_\star \sum_{|j|\geqslant m} v_j \right\|_\infty + \left\| \sum_{|k|<m} J^\star_k \sum_{\substack{k_1+k_2=k \\ |k_2|\geqslant m}} \mathbb{A}_{k_1} v_{k_2} \right\|_\infty + \left\| \sum_{|k|<m} J^\star_k (\mathcal{E} * v^{(I)})_k \right\|_\infty \\ &\leqslant \left\| J^\star_\star \sum_{|j|\geqslant m} v_j \right\|_\infty + \left\| \sum_{|j|\geqslant m} \left(\sum_{|k|<m} J^\star_k \mathbb{A}_{k-j} \right) v_j \right\|_\infty + \left\| \sum_{|k|<m} J^\star_k (\mathcal{E} * v^{(I)})_k \right\|_\infty.\end{aligned}$$

Applying repeatedly Lemma 3.2 and Corollary 3.1, we conclude that

$$\|[JC]_\star\|_\infty \leqslant \frac{\|J^\star_\star\|_\infty}{\nu^m} \|v\|_{1,\nu} + \left\| \sum_{|k|<m} J^\star_k \mathbb{A}_{k-j} \right\|_{\infty,\nu^{-1}} \|v\|_{1,\nu} + \|J^\star\|_{\infty,\nu^{-1}} r_A \|v\|_{1,\nu}.$$

Note that $J^\star_k \mathbb{A}_{k-j} \neq 0$ only for $m \leqslant |j| \leqslant m + M_A - 2$, therefore,

$$\left\| \sum_{|k|<m} J^\star_k \mathbb{A}_{k-j} \right\|_{\infty,\nu^{-1}} = \max_{m\leqslant|j|\leqslant m+M_A-2} \frac{1}{\nu^{|j|}} \left\| \sum_{|k|<m} J^\star_k \mathbb{A}_{k-j} \right\|_\infty.$$

We now aim at computing $\|JC\|_{1,\nu}$. We see JC as a linear functional of v. According to the definition of $c_{k,1}$, we have for $|k| < m$

$$\begin{aligned}(JC(v))_k &= \sum_{|j|<m} J^k_j C_j(v) + J^k_\star C_\star(v) = \sum_{|j|<m} J^k_j \left(\sum_{\substack{a+b=j \\ |b|\geqslant m}} \mathcal{A}_a v_b \right) + J^k_\star \sum_{|b|\geqslant m} v_b \\ &= \sum_{|b|\geqslant m} \left(\sum_{|j|<m} J^k_j \mathcal{A}_{j-b} + J^k_\star \right) v_b,\end{aligned}$$

and for $|k| \geqslant m$

$$(JC(v))_k = \Lambda_k^{-1} C_k(v) = \Lambda_k^{-1} \sum_{\substack{a+b=k \\ b \neq k}} \mathcal{A}_a v_b = \sum_{b \neq k} \Lambda_k^{-1} \mathcal{A}_{k-b} v_b.$$

The ν-norm of $JC(v)$ is estimated by $\|JC(v)\|_{1,\nu} \leqslant \|\!|JC|\!\| \|v\|_{1,\nu}$. The operator norm $\|\!|JC|\!\|$ is bounded as shown in Lemma 3.3:

$$\|\!|JC|\!\| \leqslant \sup_{b \in \mathbb{Z}} \frac{1}{\nu^{|b|}} \sum_{k \in \mathbb{Z}} \|(JC)_b^k\|_\infty \nu^{|k|},$$

where

$$(JC)_b^k = \begin{cases} \displaystyle\sum_{|j|<m} J_j^k \mathcal{A}_{j-b} + J_\star^k, & |k| < m,\ |b| \geqslant m, \\ \Lambda_k^{-1} \mathcal{A}_{k-b}, & |k| \geqslant m,\ b \neq k, \\ 0, & \text{otherwise.} \end{cases}$$

In particular, the matrix JC is zero in the inner square $|b|, |k| < m$. We can then decompose

$$\|\!|JC|\!\| \leqslant \sup\{Q_1, Q_2\},$$

where Q_1 is the sup over the column $|b| < m$ and Q_2 is the sup over the columns $|b| \geqslant m$. Then

$$\begin{aligned} Q_1 = \sup_{|b|<m} Q_1(b) &= \sup_{|b|<m} \left\{ \frac{1}{\nu^{|b|}} \sum_{|k| \geqslant m} \|\Lambda_k^{-1} \mathcal{A}_{k-b}\|_\infty \nu^{|k|} \right\}, \\ Q_2 = \sup_{|b| \geqslant m} Q_2(b) &= \sup_{|b| \geqslant m} \frac{1}{\nu^{|b|}} \Bigg\{ \sum_{|k| \geqslant m, k \neq b} \|\Lambda_k^{-1} \mathcal{A}_{k-b}\|_\infty \nu^{|k|} \\ &\qquad + \sum_{|k|<m} \left\| \sum_{|j|<m} J_j^k \mathcal{A}_{j-b} + J_\star^k \right\|_\infty \nu^{|k|} \Bigg\}. \end{aligned}$$

We have

$$\begin{aligned} Q_1(b) &\leqslant \frac{1}{\nu^{|b|}} \frac{1}{\lambda_m} \left(\sum_{|k| \geqslant m} \|\mathbb{A}_{k-b}\|_\infty \nu^{|k|} + \sum_{|k| \geqslant m} \|\mathcal{E}_{k-b}\|_\infty \nu^{|k|} \right) \\ &\leqslant \frac{1}{\nu^{|b|}} \frac{1}{\lambda_m} \sum_{\substack{|k| \geqslant m \\ |k-b|<M_A}} \|\mathbb{A}_{k-b}\|_\infty \nu^{|k|} + \frac{1}{\Lambda_m} \sum_{|k| \geqslant m} \|\mathcal{E}_{k-b}\|_\infty \nu^{|k-b|} \\ &\leqslant \frac{1}{\nu^{|b|}} \frac{1}{\lambda_m} \sum_{\substack{|k| \geqslant m \\ |k-b|<M_A}} \|\mathbb{A}_{k-b}\|_\infty \nu^{|k|} + \frac{r_A}{\lambda_m}. \end{aligned}$$

As for the terms $Q_2(b)$, we first note that the term $\mathcal{A}_0$ never appears, therefore we introduce $\widetilde{\mathcal{A}}$ to be the same as $\mathcal{A}$ but $\widetilde{\mathcal{A}}_0 = 0$ and we forget about the condition $k \neq b$:

$$Q_2(b) \leqslant \frac{1}{\nu^{|b|}}\bigg(\frac{1}{\Lambda_m}\sum_{|k|\geqslant m}\|\widetilde{A}_{k-b}\|_\infty \nu^{|k|} + \sum_{|k|<m}\|J^k_\star\|_\infty \nu^{|k|} + \sum_{|k|<m}\bigg\|\sum_{|j|<m} J^k_j \mathcal{A}_{j-b}\bigg\|_\infty \nu^{|k|}\bigg).$$

Write $\mathcal{A} = \mathbb{A} + \mathcal{E}$ and denote $\mathcal{F}_j = \mathcal{E}_{j-b}$. Thus $\|\mathcal{F}\|_{1,\nu} \leqslant \nu^{|b|} r_A$ and

$$\sum_{|k|<m}\bigg\|\sum_{|j|<m} J^k_j \mathcal{A}_{j-b}\bigg\|_\infty \nu^{|k|} \leqslant \sum_{|k|<m}\bigg\|\sum_{|j|<m} J^k_j \mathbb{A}_{j-b}\bigg\|_\infty \nu^{|k|} + \sum_{|k|<m}\bigg\|\sum_{|j|<m} J^k_j \mathcal{E}_{j-b}\bigg\|_\infty \nu^{|k|}.$$

By Lemma 3.2, the last term is bounded by

$$\sum_{|k|<m}\bigg\|\sum_{|j|<m} J^k_j \mathcal{F}_j\bigg\|_\infty \nu^{|k|} \leqslant \sum_{|k|<m}\|J^k\|_{\infty,\nu^{-1}}\|\mathcal{F}\|_{1,\nu}\nu^{|k|} \leqslant \nu^{|b|} r_A \sum_{|k|<m}\|J^k\|_{\infty,\nu^{-1}}\nu^{|k|}$$

which, when collecting all the contributions, results in

$$Q_2(b) \leqslant \frac{1}{\nu^{|b|}}\bigg(\frac{1}{\Lambda_m}\sum_{|k|\geqslant m}\|\mathbb{A}_{k-b}\|_\infty \nu^{|k|} + \sum_{|k|<m}\bigg\|\sum_{|j|<m} J^k_j \mathbb{A}_{j-b}\bigg\|_\infty \nu^{|k|}\bigg) + \frac{1}{\nu^{|b|}}\|J_\star\|_{1,\nu} + \frac{1}{\lambda_m} r_A + r_A \sum_{|k|<m}\|J^k\|_{\infty,\nu^{-1}}\nu^{|k|}.$$

Since we cannot compute $Q_2(b)$ for all $|b| \geqslant m$, we need a uniform estimate for $|b|$ large enough. We set the threshold at $|b| = m + M_A$ so that the second of the above sums vanishes.

For $|b| \geqslant m + M_A$,

$$Q_2(b) \leqslant \frac{\|\widetilde{\mathbb{A}}\|_{1,\nu} + r_A}{\Lambda_m} + \frac{1}{\nu^{m+M_A}}\|J_\star\|_{1,\nu} + r_A \sum_{|k|<m}\|J^k\|_{\infty,\nu^{-1}}\nu^{|k|}.$$

4.2.3. Compute $Z^{(2)}$. From (4.5), $Z^{(2)}$ is defined so that

$$Z^{(2)} \geqslant \sup_{\|u\|_X\leqslant 1, \|v\|_X\leqslant 1} \max\{\|[JD]_\star\|_\infty, \|\{(JD)_k\}_{k\in\mathbb{Z}}\|_{1,\nu}\},$$

where $D = \{c_{\star,2}, \{c_{k,2}\}_k\}$. We compute $[JD]_\star = J^\star_\star D_\star + \sum_{|k|<m} J^\star_k D_k$, hence, for $\|u\|_X, \|v\|_X \leqslant 1$,

$$\|[JD]_\star\|_\infty \leqslant 2\bigg\|\sum_{|k|<m} J^\star_k (u_k v_\star)\bigg\|_\infty \leqslant 2\|J^\star\|_{\infty,\nu^{-1}}\|u\|_{1,\nu}\|v_\star\|_\infty \leqslant 2\|J^\star\|_{\infty,\nu^{-1}}.$$

Since $D_\star = c_{\star,2} = 0$, it follows that

$$\|JD\|_{1,\nu} \leqslant |\!|\!| J |\!|\!| \, \|D\|_{1,\nu},$$

where $|\!|\!| J |\!|\!|$ is bounded as in Lemma 3.4 and $\|D\|_{1,\nu} \leqslant 2$.

5. Application and computational results

As mentioned in the introduction, the fundamental matrix solution plays an important role in dynamical system theory. The application we are most interested in is the following: let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be a vector field and consider the equation $\dot{x} = f(x)$. Suppose Γ is a hyperbolic τ-periodic orbit parameterized by $\gamma(t)$. System (1.1) with $A(t) = Df(\gamma(t))$ is the linearized (or variational) system along $\gamma(t)$ and, in this context, the matrix $\Phi(\tau)$ is called the monodromy matrix. The spectral data of the monodromy matrix encode the stability properties of the orbit such as the directions tangent to the stable/unstable manifold at the point $\gamma(0)$. As explained in [5], the Floquet normal form decomposition of the fundamental matrix solution provides the spectral data of the monodromy matrix and, more interestingly, allows a continuous parameterization of the normal stable/unstable bundles of Γ, or equivalently the tangent bundles of the stable/unstable manifolds along Γ.

As a toy model, let us consider the Lorenz system, given by the vector field

$$f(x,y,z) = \begin{pmatrix} -\sigma x + \sigma y \\ \varrho x - y - xz \\ -\beta z + xy \end{pmatrix}, \quad s, \beta, \varrho \in \mathbb{R}, \tag{5.1}$$

where $\beta = 8/3$, $\sigma = 10$, and $\varrho = 28$. For this set of parameters it is well known that the Lorenz system is chaotic. Suppose that a τ-periodic solution in the form $\gamma(t) = \sum\limits_{k\in\mathbb{Z}} \gamma_k \mathrm{e}^{\mathrm{i}k2\pi/\tau t}$ has been rigorously enclosed in the space of analytic functions, see for instance [9]. For example, a periodic solution $\gamma(t)$ has been proved to exist according to the following bounds:

$$\begin{gathered} r_\gamma = 2.109574 \cdot 10^{-10}, \qquad \nu_\gamma = 1.21, \\ |\omega - 4.031165685315| < r_\gamma, \quad \|\gamma - \overline{\gamma}\|_{\nu_\gamma} \leqslant r_\gamma, \quad \overline{\gamma}_k = 0 \quad \forall\, |k| > 60, \end{gathered}$$

where $\omega = 2\pi/\tau$ and $\gamma = \{\gamma_k\}_{k\in\mathbb{Z}}$.

In order to determine the Floquet normal form of the fundamental matrix solution of the linearization around $\gamma(t)$, we solve system (2.4) (2.5) with

$$\mathcal{A}_0 = \begin{pmatrix} -\sigma & \sigma & 0 \\ \varrho - \gamma_0^{(3)} & -1 & -\gamma_0^{(1)} \\ \gamma_0^{(2)} & \gamma_0^{(1)} & -\beta \end{pmatrix}, \quad \mathcal{A}_{2k} = \begin{pmatrix} 0 & 0 & 0 \\ -\gamma_k^{(3)} & 0 & -\gamma_k^{(1)} \\ \gamma_k^{(2)} & \gamma_k^{(1)} & 0 \end{pmatrix}, \quad \mathcal{A}_{2k+1} = 0.$$

Note that the function $A(t)$ is expanded to the 2τ-periodic basis function, hence the odd Fourier coefficients are set to zero.

Referring to Section 2.1, we set $M_A = 119$, $\nu = \sqrt{\nu_\gamma} = 1.1$, $r_A = 6.328 \cdot 10^{-10}$, define $\mathbb{A}_k$ as before with $\overline{\gamma}$ in place of γ. This choice ensures that $\|\mathcal{A} - \mathbb{A}\|_{1,\nu} \leqslant r_A$.

Then we choose the finite-dimensional parameter $m = 100$ and compute numerically an approximate solution $\bar{x} = (\overline{R}, \overline{Q})$. (That could be done by numerically integrating first system (1.2) for one period and then integrating system (1.2) and extracting a first guess of the matrices R and $\mathcal{Q}$. Subsequently, these data are optimized by means of a Newton scheme. See [5] for further details). The rigorous computation of the estimates described in Section 4 produces the bounds

$$Y = 2.5221 \cdot 10^{-6}, \quad Z^{(0)} = 1.4546 \cdot 10^{-6}, \quad Z^{(1)} = 0.531124, \quad Z^{(2)} = 161.416,$$

and the radii polynomial $p(r)$ is negative for any $r \in I = [5.3891 \cdot 10^{-6}, 0.002899]$.

In conclusion, choosing $r = 5.39 \cdot 10^{-6} \in I$, the real Floquet normal form decomposition of the fundamental matrix solution Φ of the linearization around Γ is $\Phi(t) = Q(t)\mathrm{e}^{Rt}$, where

- ▷ $\|R - \overline{R}\|_\infty \leqslant r$,
- ▷ the matrix function $Q(t)$ satisfies the expansion (2.3) with $\|\mathcal{Q} - \overline{\mathcal{Q}}\|_{1,\nu} \leqslant r$.

In particular, any component $Q_{ij}(t)$ is an analytic function and extends to a 2τ-periodic, analytic function on a complex strip $\mathbb{S}_\omega = \{z \in \mathbb{C}\colon\ z = a + \mathrm{i}b \text{ and } |b| < \omega\}$, where the width ω is related to the decay rate ν.

We repeat the computation for different orbits on the bifurcation branch which, from the Hopf bifurcation at $\varrho \sim 24.8$, moves towards the homoclinic point at $\varrho \sim 13.9$. The results are reported in the table on the right and are compared with the results obtained by computing the enclosure in the C^k category [5] (left table). The parameter s refers to the norm $\|\mathcal{Q}\|_s = \sup\{\|\mathcal{Q}_0\|_\infty, \|\mathcal{Q}_k\|_\infty |k|^s\}$ used in the C^k enclosure. We recall once more that the s-norm implies that the solution is C^s-smooth.

C^k category				
ϱ	m_γ	m	s	r
22	30	50	3	$3.51 \cdot 10^{-9}$
20	30	50	2	$3.60 \cdot 10^{-10}$
17.32	60	90	2	$3.91 \cdot 10^{-9}$
15	60	140	2	$2.10 \cdot 10^{-8}$

Analytic category				
ϱ	m_γ	m	ν	r
22	30	50	1.5	$4.95 \cdot 10^{-8}$
20	30	50	1.2	$7.28 \cdot 10^{-9}$
17.32	50	60	1.2	$3.34 \cdot 10^{-7}$
15	50	80	1.1	$2.52 \cdot 10^{-5}$

As expected, both the methods require a larger finite-dimensional projection as the orbit approaches the homoclinic point. However, it seems that the analytic enclosure can be still achieved with a smaller number of modes compared to the C^k

enclosure. On the other hand, the algorithm in the C^k class seems to yield sharper enclosures. To ensure mathematical rigor, all the computations are performed in MATLAB equipped with the INTLAB package [12]. INTLAB accumulates all possible floating point rounding errors using interval arithmetics and returns validated interval enclosures of all numerical results.

The results of these preliminary comparisons should not be read as definitive, and are provided only in order to illustrate that both C^k and analytic arguments produce good results with reasonable computational costs. At present when one method outperforms the other it is not always clear to us whether to thank the theory or blame the implementation. Indeed, moving the methods discussed here beyond the "proof of concept" phase is a topic of ongoing research. However, it is perhaps more important to remark that these are complementary and not competing methods. When solutions of the differential equation are a priori analytic then only the analytic formulation will provide quantitative information about analytic properties such as domain of analyticity and exponential rate of decay for the Fourier coefficients. On the other hand, if the solution is known a priori to be C^k and not analytic then the analytic formulation is doomed to fail, and we must pursue the proof in a space of algebraic decay. Taken together the methods facilitate the study of a larger class of questions than could be addressed with either method singly.

References

[1] *X. Cabré, E. Fontich, R. de la Llave*: The parameterization method for invariant manifolds I: Manifolds associated to non-resonant subspaces. Indiana Univ. Math. J. *52* (2003), 283–328. zbl MR

[2] *X. Cabré, E. Fontich, R. de la Llave*: The parameterization method for invariant manifolds II: Regularity with respect to parameters. Indiana Univ. Math. J. *52* (2003), 329–360. zbl MR

[3] *X. Cabré, E. Fontich, R. de la Llave*: The parameterization method for invariant manifolds III: Overview and applications. J. Differ. Equations *218* (2005), 444–515. zbl MR

[4] *R. Castelli, J.-P. Lessard*: A method to rigorously enclose eigenpairs of complex interval matrices. Internat. Conf. Appl. Math. In Honor of the 70th Birthday of K. Segeth. Academy of Sciences of the Czech Republic, Institute of Mathematics, Prague, 2013, pp. 21–31. MR

[5] *R. Castelli, J.-P. Lessard*: Rigorous numerics in Floquet theory: computing stable and unstable bundles of periodic orbits. SIAM J. Appl. Dyn. Syst. (electronic only) *12* (2013), 204–245. zbl MR

[6] *R. Castelli, J.-P. Lessard, J. D. Mireles James*: Parameterization of invariant manifolds for periodic orbits I: Efficient numerics via the Floquet normal form. SIAM J. Appl. Dyn. Syst. (electronic only) *14* (2015), 132–167. zbl MR

[7] *S. Day, J.-P. Lessard, K. Mischaikow*: Validated continuation for equilibria of PDEs. SIAM J. Numer. Anal. *45* (2007), 1398–1424. zbl MR

[8] *M. Gameiro, J.-P. Lessard*: Analytic estimates and rigorous continuation for equilibria of higher-dimensional PDEs. J. Differ. Equations *249* (2010), 2237–2268. zbl MR

[9] *A. Hungria, J.-P. Lessard, J. D. Mireles James*: Rigorous numerics for analytic solutions of differential equations: the radii polynomial approach. To appear in Math. Comput. (2015).

[10] *J.-P. Lessard, J. D. M. James, C. Reinhardt*: Computer assisted proof of transverse saddle-to-saddle connecting orbits for first order vector fields. J. Dyn. Differ. Equations *26* (2014), 267–313. zbl MR

[11] *J. D. Mireles James, K. Mischaikow*: Rigorous a posteriori computation of (un)stable manifolds and connecting orbits for analytic maps. SIAM J. Appl. Dyn. Syst. (electronic only) *12* (2013), 957–1006. zbl MR

[12] *S. M. Rump*: INTLAB—INTerval LABoratory. Developments in Reliable Computing (T. Csendes, ed.). SCAN-98 conference, Budapest. Kluwer Academic Publishers, Dordrecht, 1999, pp. 77–104, `http://www.ti3.tu-harburg.de/rump/`. zbl

[13] *V. A. Yakubovich, V. M. Starzhinskij*: Linear Differential Equations with Periodic Coefficients, Vol. 1, 2. Wiley, New York, Halsted, Jerusalem, 1975. zbl MR

Authors' addresses: *Roberto Castelli*, Faculty of Sciences, Department of Mathematics, VU University Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands, e-mail: `r.castelli@vu.nl`; *Jean-Philippe Lessard*, Département de Mathématiques et de Statistique, Université Laval, 1045 avenue de la Médecine, Québec, G1V0A6, Canada, e-mail: `jean-philippe.lessard@mat.ulaval.ca`; *Jason D. Mireles James*, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA, e-mail: `jmirelesjames@fau.edu`.